

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 25, 2003, 21:56:10 ; Search time 4670.87 Seconds

(without alignments)  
11088.207 Million cell updates/sec

Title: US-09-717-321A-2

Perfect score: 1266

Sequence: 1 ttttttttttttttcca.....gagcaagaatagggtttt 1266

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 2888711 seqs, 2045481386 residues

Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

GenBank1:  
1: gb\_ba:\*  
2: gb\_bt:\*  
3: gb\_in:\*  
4: gb\_lm:\*  
5: gb\_ov:\*  
6: gb\_pat:\*  
7: gb\_ph:\*  
8: gb\_pl:\*  
9: gb\_pr:\*  
10: gb\_ro:\*  
11: gb\_sts:\*  
12: gb\_sy:\*  
13: gb\_un:\*  
14: gb\_vl:\*  
15: em\_ba:\*  
16: em\_fun:\*  
17: em\_hum:\*  
18: em\_in:\*  
19: em\_mu:\*  
20: em\_com:\*  
21: em\_or:\*  
22: em\_ov:\*  
23: em\_pat:\*  
24: em\_ph:\*  
25: em\_pl:\*  
26: em\_ro:\*  
27: em\_sts:\*  
28: em\_un:\*  
29: em\_vl:\*  
30: em\_htg\_hum:\*  
31: em\_htg\_inv:\*  
32: em\_htg\_mus:\*  
33: em\_htg\_other:\*  
34: em\_htg\_pln:\*  
35: em\_htg\_rnd:\*  
36: em\_htg\_mam:\*  
37: em\_htg\_vrt:\*  
38: em\_sy:\*  
39: em\_htgo\_hum:\*  
40: em\_htgo\_mus:\*  
41: em\_htgo\_other:\*

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1266	100.0	1266	6	AX163738
2	1227.4	97.0	217700	2	AX106124
3	1017	80.3	1017	6	AX163751
4	878.2	69.4	2319	10	BC003828
5	868.2	68.6	2281	10	BC051053
6	862	68.1	269081	2	AC068493
7	445.6	35.2	2302	9	BC050687
8	441.8	34.9	28567	9	HSX132695
9	441.8	34.9	212827	9	AC009412
10	430.6	34.0	2315	9	AK054993
11	420.2	33.2	1232	6	AX408985
12	420.2	33.2	1232	9	HUMPO2ST9
13	408.2	32.2	185745	2	AC102775
14	408.2	32.2	202342	10	AL583884
15	369.4	29.2	192498	2	AC105379
16	369.4	29.2	230015	2	AC132602
17	348	27.5	348	6	AX163753
18	337.2	26.6	179685	10	AC126556
19	309.6	24.5	137625	9	AC104663
20	306.4	24.2	5544	9	AF542527
21	290.6	23.0	1022	6	AX163752
22	250.4	19.8	87616	2	AC139405
23	194.8	15.4	202565	9	AL354696
24	193.2	15.3	605	6	AX341407
25	187.8	14.8	352	6	AX163754
26	175.6	13.9	240973	2	AC123247
27	175.6	13.9	241048	2	AC129624
28	170.6	13.5	262	11	G31709
29	154	12.2	455	11	G26995
30	131.4	10.4	228121	2	AC133022
31	131.4	10.4	239768	2	AC112582
32	131.4	10.4	245468	2	AC130746
33	123	9.7	123	6	AX163737
34	118	9.3	137625	9	AC104663
35	116.4	9.2	110816	9	AC002404
36	112.4	8.9	64781	2	AC102025
37	110.6	8.7	101584	9	CNS01DS5
38	110.6	8.7	155943	9	AC012364
39	105.2	8.3	3544	9	AF542527
40	98.4	7.8	170839	2	AC133467
41	86.6	6.8	180303	9	AL672045
42	86.6	6.8	201012	2	AC021189
43	85.6	6.8	495	6	AX341256
44	84.6	6.7	174316	2	AC022647
45	84.6	6.7	217249	9	AC009902

#### ALIGNMENTS

RESULT 1  
LOCUS AX163738 1266 bp DNA linear PAT 22-JUN-2001  
DEFINITION Sequence 2 from Patent WO0138579.  
ACCESSION AX163738  
VERSION AX163738.1 GI:1544844  
KEYWORDS  
SOURCE  
ORGANISM Rattus norvegicus (Norway rat)  
Rattus norvegicus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;  
Rattus.  
REFERENCE  
1 Gould-Rothberg,B.E., Dippio,V.A., Ramsch,T.M. and Gerwein,R.W.  
Method of Identifying toxic agents using nsaid-induced differential



Milosavljevic, A., Miner, G., Minja, E., Montemayor, J., Moore, S.,  
 Morgan, M., Morris, K., Morris, S., Muidasa, M., Murphy, M., Nair, L.,  
 Nankevis, C., Neal, D., Newton, N., Nguyen, N., Norris, S.,  
 Nwokileh, O., Okunolu, G., Olampunsaogon, A., Pal, S., Parks, K.,  
 Pasternak, S., Paul, H., Perez, A., Perez, L., Pflankoch, C.,  
 Plopper, F., Polidexter, A., Popovic, D., Primus, E., Pu, L., L.,  
 Puzo, M., Quiroz, J., Rachlin, E., Reeves, K., Regier, M. A., Reigh, R.,  
 Rell, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F.,  
 Rives, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, J.,  
 Sanders, W., Savary, G., Scherer, S., Scott, G., Shatman, S., Shen, H.,  
 Shetty, J., Shvartsbeyn, A., Slisson, I., Sitter, C. D., Smajs, D.,  
 Sneed, A., Sodergren, E., Song, X.-Z., Sorelle, R., Sosa, J.,  
 Steimle, M., Strong, R., Sutton, A., Swalek, A., Taber, P., Taylor, C.,  
 Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Usmani, K.,  
 Valas, R., Vera, V., Villasana, D., Waldron, L., Walker, B., Wang, J.,  
 Wang, O., Wang, S., Warren, J., Warren, R., Wei, X., White, F.,  
 Williams, G., Willson, R., Wleczek, R., Wooden, H., Worley, K.,  
 Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V.,  
 Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von  
 Niederhausern, A., Weiss, R., Smith, D. R., Holt, R. A., Smith, H. O.,  
 Weinstock, G. and Gibbs, R. A.  
 Direct Submission  
 Unpublished  
 2 (bases 1 to 217700)  
 Worley, K. C.  
 Direct Submission  
 Submitted (12-JAN-2002) Human Genome Sequencing Center, Department  
 of Molecular and Human Genetics, Baylor College of Medicine, One  
 Baylor Plaza, Houston, TX 77030, USA  
 3 (bases 1 to 217700)  
 Rat Genome Sequencing Consortium.  
 Direct Submission  
 Submitted (10-MAY-2003) Human Genome Sequencing Center, Department  
 of Molecular and Human Genetics, Baylor College of Medicine, One  
 Baylor Plaza, Houston, TX 77030, USA  
 On May 10, 2003 this sequence version replaced gi:25094619.  
 The sequence in this assembly is a combination of BAC based reads  
 and whole genome shotgun sequencing reads assembled using Atlas  
 (<http://www.hgsc.bcm.tmc.edu/projects/rat/>). Each contig described  
 in the feature table below represents a scaffold in the Atlas  
 assembly (a 'contig-scaffold'). Within each contig-scaffold,  
 individual sequence contigs are ordered and oriented, and separated  
 by sized gaps filled with Ns to the estimated size. The sequence  
 may extend beyond the ends of the clone and there may be sequence  
 contigs within a contig-scaffold that consist entirely of whole  
 genome shotgun sequence reads. Both end sequences and whole genome  
 shotgun sequence only contigs will be indicated in the feature  
 table.  
 ----- Genome Center  
 Center: Baylor College of Medicine  
 Center code: BCM  
 Web site: <http://www.hgsc.bcm.tmc.edu/>  
 Contact: hgsc-help@bcm.tmc.edu  
 ----- Project Information  
 Center project name: GHLL  
 Center clone name: CH230-119E11  
 ----- Summary Statistics  
 Assembly program: Atlas 3.0:  
 Consensus quality: 185725 bases at least Q40  
 Consensus quality: 190140 bases at least Q30  
 Consensus quality: 192842 bases at least Q20  
 Estimated insert size: 194758; sum-of-contigs estimation  
 Quality coverage: 5x in Q20 bases; sum-of-contigs estimation  
 -----  
 \* NOTE: Estimated insert size may differ from sequence length  
 \* (see [http://www.hgsc.bcm.tmc.edu/docs/genbank\\_draft\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html)).  
 \* NOTE: This is a "working draft" sequence. It currently  
 \* consists of 11 contigs. The true order of the pieces  
 \* is not known and their order in this sequence record is  
 \* arbitrary. Gaps between the contigs are represented as  
 \* runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence  
 \* as soon as it is available and the accession number will

\* be preserved.  
 \* 1 109797: contig of 109797 bp in length  
 \* 109798 109897: gap of unknown length  
 \* 109898 166934: contig of 57037 bp in length  
 \* 166935 167034: gap of unknown length  
 \* 167035 176384: contig of 9350 bp in length  
 \* 176385 176484: gap of unknown length  
 \* 176485 206327: contig of 29843 bp in length  
 \* 206328 206427: gap of unknown length  
 \* 206428 207461: contig of 1034 bp in length  
 \* 207462 207561: gap of unknown length  
 \* 207562 208607: contig of 1046 bp in length  
 \* 208608 208707: gap of unknown length  
 \* 208708 209908: contig of 1201 bp in length  
 \* 209909 210008: gap of unknown length  
 \* 210009 211583: contig of 1575 bp in length  
 \* 211584 211683: gap of unknown length  
 \* 211684 213649: contig of 1966 bp in length  
 \* 213650 213749: gap of unknown length  
 \* 213750 215115: contig of 1366 bp in length  
 \* 215116 215215: gap of unknown length  
 \* 215216 217700: contig of 2485 bp in length.  
 FEATURES  
 source  
 1. 217700  
 /organism="Rattus norvegicus"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:10116"  
 /clone="CH230-119E11"  
 BASE COUNT 51384 a 45640 c 45954 g 51476 t 23246 others  
 ORIGIN  
 Query Match 97.0%; Score 1227.4; DB 2; Length 217700;  
 Best Local Similarity 98.7%; Pred. No. 1,6e-312;  
 Matches 1237; Conservative 0; Mismatches 16; Indels 0; Gaps 0;  
 14 TTTTCAGTCCCAAGACATTTTATGATTCAAGATTATTAAGTCA 73  
 |||||||||  
 64694 TTGTCAAGTCCCAAGACCAAGCATGATTTATGATTCAGATTTATTAAGTCA 64635  
 |||||||||  
 74 TACATGCAAAACATCTCTAATCAGATTCAGAAAGTCAATGTAAGTCAAGTCA 133  
 TACATGCAAAACATCTCTAATCAGATTCAGAAAGTCAATGTAAGTCAAGTCAAGTCA 64575  
 |||||||||  
 134 TTCTGCACTGTCATTTAAAAAGTTGTTCTAGTGTGCAAGGCCAACACTGTGT 193  
 |||||||||  
 64574 TTCTGTAAGTGTCAATTTAAAAAGTTGTTCTAGTGTGCAAGGCCAACACTGTGT 64515  
 |||||||||  
 194 CTTCGCACTGTAAGTGTGTAAGACAGGCGTTAGCACTAGCGCTTACAGAACCTCAG 253  
 |||||||||  
 64514 CTTCGCACTGTAAGTGTGTAAGACAGGCGTTAGCACTAGCGCTTACAGAACCTCAG 64455  
 |||||||||  
 254 AGACCCAAAGTACCGGAAGCATGTCGCCGTGCTAGAGTGTAGAGGGGGCGCATCA 313  
 AGACCCAAAGTACCGGAAGCATGTCGCCGTGCTAGAGTGTAGAGGGGGCGCATCA 64395  
 |||||||||  
 64454 AGACCCAAAGTACCGGAAGCATGTCGCCGTGCTAGAGTGTAGAGGGGGCGCATCA 64395  
 |||||||||  
 314 ATCATGATGACAGTGTGTACTCTGCAAGACAGATGTTTACAGATATCTAAATAGT 373  
 |||||||||  
 64394 ATCATGATGACAGTGTGTACTCTGCAAGACAGATGTTTACAGATATCTAAATAGT 64335  
 |||||||||  
 374 TTTAAAACTGTAAAGCCGACAGCTGATTTCTACACCAGTTACTAGAAAAAGGA 433  
 TTTAAAACTGTAAAGCCGACAGCTGATTTCTACACCAGTTACTAGAAAAAGGA 64275  
 |||||||||  
 64334 TTTAAAACTGTAAAGCCGACAGCTGATTTCTACACCAGTTACTAGAAAAAGGA 64275  
 |||||||||  
 434 AGCATAGTACAGCTAGTAAAGGAAGTGAAGAAAGCAAGCACTTCTACTATACCA 493  
 AGCATAGTACAGCTAGTAAAGGAAGTGAAGAAAGCAAGCACTTCTACTATACCA 64215  
 |||||||||  
 64214 AGCATAGTACAGCTAGTAAAGGAAGTGAAGAAAGCAAGCACTTCTACTATACCA 64215  
 |||||||||  
 494 AAAAATCTCCGAATCATTTATCAGAAAGATCTTATAGTACAGTGTGCTGCTGT 553  
 AAAAATCTCCGAATCATTTATCAGAAAGATCTTATAGTACAGTGTGCTGCTGT 64155  
 |||||||||  
 64214 AAAAATCTCCGAATCATTTATCAGAAAGATCTTATAGTACAGTGTGCTGCTGT 64155  
 |||||||||  
 554 TAAGAGGGGGTCTTAAAGAAAGCACTTGTAGTGAACAGTGTGAGATGGCAGTT 613  
 |||||||||

Db 64154 TAAAGAGGGGCTCTAAGAAAAAGCACTGCTAAGTTACACTGTGAGAGTCCAGTT 64095  
QY 614 TAAATATGACTCAACGCCCATCTGTGGAGAGGACAGCAGAGGGGAGGAGGCTCAAGAG 673  
Db 64094 TAAATATGACTCAACGCCCATCTGTGGAGAGGACAGCAGAGGGGAGGAGGCTCAAGAG 64035  
QY 674 AGACACTGATTAAGATGGCCATTTGTCACTCTACTGTTGACAGAAATTAACCTTAATA 733  
Db 64034 AGACACTGATTAAGATGGCCATTTGTCACTCTACTGTTGACAGAAATTAACCTTAATA 63975  
QY 734 GCTTACCGGTGACCTTTTATTTAGTTGATTAATACCTAATGATTAATTAATTA 793  
Db 63974 GCTTACCGGTGACCTTTTATTTAGTTGATTAATACCTAATGATTAATTAATTA 63915  
QY 794 ATCTTACTTCAATTAATTAAGTCAAAATATCTGTCTCTCTCTTGTGATGACGTCTTTCACA 853  
Db 63914 ATCTTACTTCAATTAATTAAGTCAAAATATCTGTCTCTCTCTTGTGATGACGTCTTTCACA 63855  
QY 854 CACTCACCAGCAGACCCACGACGCTAGAAACAGAAATTAATCTTGAAGGACACAGAGAG 913  
Db 63854 CACTCACCAGCAGACCCACGACGCTAGAAACAGAAATTAATCTTGAAGGACACAGAGAG 63795  
QY 914 CCAGATCTCTGTTCAAAAGCCTGAGAAAGCCGCTGACGTGATTTTGAAGAACTCACTAT 973  
Db 63794 CCAGATCTCTGTTCAAAAGCCTGAGAAAGCCGCTGACGTGATTTTGAAGAACTCACTAT 63735  
QY 974 GAAATCAAAAGACAGAGCTGTTCACACCATCTGTGACGTACAGTAAAGTTACGTAAATGA 1033  
Db 63734 GAAATCAAAAGACAGAGCTGTTCACACCATCTGTGACGTACAGTAAAGTTACGTAAATGA 63675  
QY 1034 GCATGGGCTGATTAAGTTACAGGTGCGTTACATGGCAGCGGTGATTAAGAGGCTGTGCT 1093  
Db 63674 GCATGGGCTGATTAAGTTACAGGTGCGTTACATGGCAGCGGTGATTAAGAGGCTGTGCT 63615  
QY 1094 GTGTCAACAGGCTGTGGAGCTACGAGAGGAGTCTGCAACCCCTGAGCCAGAGAGCTGACATC 1153  
Db 63614 GTGTCAACAGGCTGTGGAGCTACGAGAGGAGTCTGCAACCCCTGAGCCAGAGAGCTGACATC 63555  
QY 1154 TTCTTAAAGACAAAGCTCTCTCAACAGCTTAAGTCTTACGTCTCTCAACAGAGCTCACT 1213  
Db 63554 TTCTTAAAGACAAAGCTCTCTCAACAGCTTAAGTCTTACGTCTCTCAACAGAGCTCACT 63495  
QY 1214 TAGTTCAACAAGATTTTGGCAATTTCTTAATCTGACAGAAATTAAGGAGATTTT 1266  
Db 63494 TAGTTCAACAAGATTTTGGCAATTTCTTAATCTGACAGAAATTAAGGAGATTTT 63442

RESULT 3  
AX163751 1017 bp DNA linear PAT 22-JUN-2001  
LOCUS AX163751  
DEFINITION Sequence 15 from Patent WO0138579.  
ACCESSION AX163751  
VERSION AX163751.1 GI:14544857  
KEYWORDS  
SOURCE Rattus norvegicus (Norway rat)  
ORGANISM Rattus norvegicus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;  
Rattus.

REFERENCE  
AUTHORS Gould-Rothberg, B.E., DiPippo, V.A., Ramsen, T.M. and Gerwehn, R.W.  
TITLE Method of identifying toxic agents using nsald-induced differential  
JOURNAL gene expression in liver  
Curegen Corporation (US)  
Patent: WO 0138579-A 15 31-MAY-2001;  
FEATURES  
source location/Qualifiers  
1. 1017  
/organism="Rattus norvegicus"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:10116"

BASE COUNT 245 a 245 c 216 g 311 t  
ORIGIN

Query Match 80.3%; Score 1017; DB 6; Length 1017;  
Best Local Similarity 100.0%; Pred. No. 3.3e-257;  
Matches 1017; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 245 GAACCTCAGACAGACCCAAAGGTACCCGAGACATGTGCCCCGTGGGTGAGAGTCTAAGGGG 304  
Db 1017 GAACCTCAGACAGACCCAAAGGTACCCGAGACATGTGCCCCGTGGGTGAGAGTCTAAGGGG 958  
QY 305 GCGGACATCAATCAGACAGAGTGTGTACTCTGGGAGAGACAGATGATGTTTCAGATATTC 364  
Db 957 GCGGACATCAATCAGACAGAGTGTGTACTCTGGGAGAGACAGATGATGTTTCAGATATTC 898  
QY 365 TAAATATGTTTAAATATGTTAAAGCCGACAGCAGTATTTCTACACCCAGTTACTAGAAA 424  
Db 897 TAAATATGTTTAAATATGTTAAAGCCGACAGCAGTATTTCTACACCCAGTTACTAGAAA 838  
QY 425 ACAGAGGAGAGCAGTGTGACGTGATTAAGGAAAGGTGAAACAGAGAGGACACTTCTACT 484  
Db 837 ACAGAGGAGAGCAGTGTGACGTGATTAAGGAAAGGTGAAACAGAGAGGACACTTCTACT 778  
QY 485 ATCTACCAAAAAATCTCCGAATGATTAATACAGAAAGATCTTATAGTACAGTCAACAT 544  
Db 777 ATCTACCAAAAAATCTCCGAATGATTAATACAGAAAGATCTTATAGTACAGTCAACAT 718  
QY 545 ATTGCTCTGTTAAGAGGGGCTCTTAAAGAAAGACCTGTGTAAGTACAGACAGTGAAGA 604  
Db 717 ATTGCTCTGTTAAGAGGGGCTCTTAAAGAAAGACCTGTGTAAGTACAGACAGTGAAGA 658  
QY 605 TGCGCAGTTTAAATATGATGATCAACGCCCATCTGGGAGAGGACAGAGGGGAGGGG 664  
Db 657 TGCGCAGTTTAAATATGATGATCAACGCCCATCTGGGAGAGGACAGAGGGGAGGGG 598  
QY 665 GCTCAAGAGAGACAGTATTAAGATGCGCCATTTGTCTACTGTTTGAACAGAAATTAAC 724  
Db 597 GCTCAAGAGAGACAGTATTAAGATGCGCCATTTGTCTACTGTTTGAACAGAAATTAAC 538  
QY 725 CGTTAAAAAGCTTTACCCGAGACATTTATTAAGTTGAATTAATCTCCATGTAACAGTGA 784  
Db 537 CGTTAAAAAGCTTTACCCGAGACATTTATTAAGTTGAATTAATCTCCATGTAACAGTGA 478  
QY 785 TGTAAATTAATCTCTACTCATATTAATGATCAAAATTAATGATGCTGCTCTTGAAGAGCTG 844  
Db 477 TGTAAATTAATCTCTACTCATATTAATGATCAAAATTAATGATGCTGCTCTTGAAGAGCTG 418  
QY 845 TGTTCACACACTCCACCCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 904  
Db 417 TGTTCACACACTCCACCCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 358  
QY 905 ACACAGAGGACAGATGCTGTTCAAAAGCCGAGAAAGCCGAGAGGCTGATTTTGAAGA 964  
Db 357 ACACAGAGGACAGATGCTGTTCAAAAGCCGAGAAAGCCGAGAGGCTGATTTTGAAGA 298  
QY 965 ACTCACTATGAATTAACAAGAGCAGAGCTGTTACACCCATCTGACGTACAGTACAAAGTT 1024  
Db 297 ACTCACTATGAATTAACAAGAGCAGAGCTGTTACACCCATCTGACGTACAGTACAAAGTT 238  
QY 1025 ACCTAATGACATGGGCTGATTAAGTTACAGGTGCTTACATGGCAGCAGTCTTAAAGA 1084  
Db 237 ACCTAATGACATGGGCTGATTAAGTTACAGGTGCTTACATGGCAGCAGTCTTAAAGA 178  
QY 1085 GCGTGTGCTGTGACAGGCTGCGGAGGTACGGGAGGGGTGACCCCTGAGCCCAAGAA 1144  
Db 177 GCGTGTGCTGTGACAGGCTGCGGAGGTACGGGAGGGGTGACCCCTGAGCCCAAGAA 118  
QY 1145 GCTGACAGTCTTCTTAAAGACAAAGTCTTCAACAGCTTATGCTTACGTCTTCAAGCAG 1204  
Db 117 GCTGACAGTCTTCTTAAAGACAAAGTCTTCAACAGCTTATGCTTACGTCTTCAAGCAG 58  
QY 1205 AACGCAACTTATGTTCAACAAAGATTTTGGCAATTTCTTAATCTGAGCAAGAAATAGGG 1261  
Db 57 AACGCAACTTATGTTCAACAAAGATTTTGGCAATTTCTTAATCTGAGCAAGAAATAGGG 1



RESULT 4  
BC003828/c  
LOCUS  
DEFINITION BC003828 2319 bp mRNA linear ROD 16-APR-2003  
MUS MUSCULUS RAS-related G3 botulinum substrate 1, mRNA (CDNA clone  
MGC:6235 IMAGE:3593957), complete cds.  
ACCESSION BC003828  
VERSION BC003828.1 GI:13277917  
KEYWORDS MGC.  
SOURCE Mus musculus (house mouse)  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Murinae; Mus.  
REFERENCE  
1 (bases 1 to 2319)  
Straussberg, R.L., Feingold, E.A., Grouse, L.H., Derge, J.G.,  
Klausner, R.D., Collins, F.S., Wagner, L., Shenmen, C.M., Schuler, G.D.,  
Altschul, S.F., Zeeberg, B., Buetow, K.H., Schaefer, C.F., Bhat, N.K.,  
Hopkins, R.F., Jordan, H., Moore, T., Max, S.I., Wang, J., Hsieh, F.,  
Diatelcon, M., Soares, M.B., Bonaldo, M.F., Casavant, T.L.,  
Scheetz, T.E., Brownstein, M.J., Usdin, T.B., Toshiyuki, S.,  
Abramson, R.D., Mullaly, S.J., Bosak, S.A., McEwan, P.J.,  
McKernan, K.J., Malek, J.A., Gunaratne, P.H., Richards, S.,  
Worley, K.C., Hale, S., Garcia, A.M., Gay, L.J., Hulyk, S.W.,  
Villalón, D.K., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A.,  
Fahey, J., Hellon, E., Kettelman, M., Madan, A., Rodriguez, S.,  
Sanchez, A., Whiting, M., Madan, A., Young, A.C., Shevchenko, Y.,  
Bouffard, G.G., Blakesley, R.W., Touchman, J.W., Green, E.D.,  
Dickson, M.C., Rodriguez, A.C., Grimwood, J., Schmutz, J., Myers, R.M.,  
Butterfield, Y.S., Krzywinski, M.I., Skalska, U., Smillius, D.E.,  
Schermer, A., Schein, J.E., Jones, S.J., and Marra, M.A.  
Generation and initial analysis of more than 15,000 full-length  
human and mouse cDNA sequences  
Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)  
22388257  
JOURNAL  
MEDLINE  
22388257  
PUBMED  
12477932  
REFERENCE  
2 (bases 1 to 2319)  
Straussberg, R.  
Direct Submission  
Submitted (28-FEB-2001) National Institutes of Health, Mammalian  
Gene Collection (MGC), Cancer Genomics Office, National Cancer  
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,  
USA  
NIH-MGC Project URL: <http://mgc.ncl.nih.gov>  
Contact: MGC help desk  
Email: [cgaps-remail.nih.gov](mailto:cgaps-remail.nih.gov)  
Tissue Procurement: Gilbert Smith, Ph.D.  
CDNA Library Preparation: Life Technologies, Inc.  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Sequencing Group at the Stanford Human Genome  
Center, Stanford University School of Medicine, Stanford, CA 94305  
Web site: <http://www-shgc.stanford.edu>  
Contact: (Dickson, Mark) [mdickpax11@stanford.edu](mailto:mdickpax11@stanford.edu)  
Dickson, M., Schmutz, J., Grimwood, J., Rodriguez, A., and Myers,  
R. M.  
Clone distribution: MGC clone distribution information can be found  
through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>  
Series: IRAX Plate: 11 Row: e Column: 22.  
FEATURES  
SOURCE  
Location/Qualifiers  
1..2319  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/strain="FVB/N"  
/db\_xref="taxon:10090"  
/clone="MGC:6235 IMAGE:3593957"  
/tissue\_type="Mammary tumor. Metallothionein-TGF alpha  
model. 10 month old virgin mouse. Taken by biopsy."  
/clone\_11b="NCI CGAP\_Mam1"  
/lab\_host="DH10B"  
/note="Vector: pCMV-SPORT6"  
1..2319  
/gene="Rac1"

CDs  
/note="synonym: D5Erd559e"  
/db\_xref="LocusID:19353"  
/db\_xref="MGI:97845"  
198..776  
/product="Rac1 protein"  
/protein\_id="AAH03828.1"  
/db\_xref="GI:13277918"  
/db\_xref="LocusID:19353"  
/translation="MQAIKCYVGGAGVKKCLLSYTNAPRGYITPVDPNSAN  
HYDKPVLGMDINGEDYDRPLPSIDIVFLCSLVSASFENRKAPEVR  
HHCPTPIIVGTGKLDRDKDIEKLEKILPTTPYDGLAMAREIGAVKYLECSAL  
TORGKLTVEDEIRAVILCPPVKKRRKCLLL"  
BASE COUNT 603 a 573 c 503 g 640 t  
ORIGIN  
Query Match 69.4%; Score 878.2; DB 10; Length 2319;  
Best local Similarity 85.1%; Pred. No. 1.5e-220;  
Matches 1116; Conservative 0; Mismatches 143; Indels 52; Gaps 10;  
1 TTTTCTTTTCTTTTCTTTTCAAGTCCAAAGCACTTTTCTTTTCTTTTATGATTCAGG 60  
|||||  
2318 TTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTT 2259  
|||||  
61 ATTTATTAATGATCATCATCAAAACATACCTGCTAATGAGCAAAAGATCAATGTGA 120  
|||||  
2258 ATTTATTAATGATCATCATCAAAACATACCTGCTAATGAGCAAAAGATCAATGTGA 2199  
|||||  
121 AAACACTCCCAATTCCTGCACTGTCATTTGAAAAAGTTGTCTAGTGTGCAAAAGGC 180  
|||||  
2198 AAACACTCCCAATTCCTGCACTGTCATTTGAAAAAGTTGTCTAGTGTGCAAAAGGC 2139  
|||||  
181 CCACACCTGTCTTCTGCGAGTATGTTAGTTGTTGACAGAGCGCTTAGCACTACGCTT 240  
|||||  
2138 CCACACCTGTCTTCTGCGAGTATGTTAGTTGTTGACAGAGCGCTTAGCACTACGCTT 2079  
|||||  
241 GACGAACTCTCAGACAGCCAAAGC-----TACCGAAGCAGTGTG 279  
|||||  
2078 GACGAACTCTCAGACAGCCAAAGC-----TACCGAAGCAGTGTG 2019  
|||||  
280 TCCGCGTGGGTAGTCTTGAAGGGGCGGCATCAATCATCATGCTGTGTTACTCTGG 339  
|||||  
2018 TCCAGCTGGGCGAGGTCTTGA---GCGACATTAATCTACATGACAGTGTGTTACTGTTT 1962  
|||||  
340 CAAGCAGTATGTTTCAAGATATCTAAATATGTTTAAATGTTAAAGCCGACAGAGT 399  
|||||  
1961 AGAAGTTCATATGTTTAAAGATATCTAAATATGTTTAAATGTTTAAAGCCGACAGAT 1902  
|||||  
400 GATTTCTACACCCAGTCTACTAGAAAACGAAGGAGCACTAGT-CAGCTAGTAAAGGAA 458  
|||||  
1901 GATTTGACACACTAGTGTGTGAAGAACTAGGAACACTAGTACTGAGTAAAGTAA 1842  
|||||  
459 GGTGAAGAACGAGACGCACTTCTACTATCTACCAAAAAAATCTCCGAA-----TGCATTA 513  
|||||  
1841 GCGGAAACAGAGCAGCCACTTCTACTAGCTACCAAAAAAATAATCTCTCGTTA 1782  
|||||  
514 TCAGAAAGATCTTATAGTACAGTCAAGCATTTCTCTGTTAAGAGGGGCTCTAAAGA 573  
|||||  
1781 TCAGAAAGATCTTATAGTACAGTCAAGCATTTCTCTGTTAAGAGGGGCTCTAAAGA 1723  
|||||  
574 AAAGCATTGCTTAAGTACCACTGTGAGATGGCCAGTTTAAATATGAGTCAACGCC 633  
|||||  
1722 AAAGCATTGCTTAAGTACCACTGTGAGATGGCCAGTTTAAATATGAGTCAACGCC 1664  
|||||  
634 CATCTGGGAGGAGACGACAGGGGGAAGGGGGCTTAAAG-----AGAGCACTG 681  
|||||  
1663 CATCTGGGAGGAGACGACAGGGGGAAGGGGGCTTAAAG-----AGAGCACTG 741  
|||||  
682 ATTAAGTGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1604  
|||||  
1603 ATTAAGTGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1544  
|||||  
742 CGTGACACTTTTATTCAGTGAATTAATCAATGATGATGATGATGATGATGATGATGATG 801  
|||||

|||||  
Db 1543 CGGACACTTATTTAGTTGAATTAAGTACTCAGTACAAATGATGTAAAGTAATCTCTAC 1484  
QY 802 TTTCATTTAGTCAAAATAGTGTCTGTCTCTTGTATGAGTGTCTTTACACACTCCAC 861  
Db 1483 TTTCATTTAGTCAAAATAGTGTCTGTCTCTTGTATGAGTGTCTTTACACACTCCAC 1426  
QY 862 CCAGCACACCCAGCAGTACAGACAGAAATCTGTCTTAAAGGCAACACAGGACCAAGTT 921  
Db 1425 CCAGCACACCCAGCAGTACAGACAGAAATCTGTCTTAAAGGCAACACAGGACCAAGTT 1366  
QY 922 CTGTCTAAAGCCGAGAAAGCCGTCAGCTGATATTTTAAAGAACTCATATGAATCA 981  
Db 1365 CTGTCTAAAGCCGAGAAAGCCGTCAGCTGATATTTTAAAGAACTCATATGAATCA 1306  
QY 982 AGAGCAGAGCTGTATACACCCATC--GTGACGTACAGTACAAAGTTACGTAATGAGCAT 1038  
Db 1305 AGAGCAGAGCTGTACACACTCATCTGTCAGTACAGTACAAAGTTACGTAATGAGCAT 1246  
QY 1039 GCGCTAATAGTTACAGTGTGCTTACATAGGAGGCTGTCTATTAAGAGGCTGTCTGTCT 1098  
Db 1245 GCGCTAATAGTTACAGTGTGCTTACATAGGAGGCTGTCTATTAAGAGGCTGTCTGTCT 1186  
QY 1099 ACAGGCTGTGAGTACAGGAGGCTGTCTGACACCCGTGAGCCGAGAAAGCTGAGTCTTCT 1158  
Db 1185 ACAGGCTGTGAGTACAGGAGGCTGTCTGACACCCGTGAGCCGAGAAAGCTGAGTCTTCT 1126  
QY 1159 AAGGACAAA--GTCTCTCAACAGCTTATGCTTACGTGTCTTCAGACACAAAGCACTTA 1215  
Db 1125 AAGGACAAAAGTCTCTCAACAGAGTATAGTCTTACGTGTCTTCAGACACAAAGCACTTA 1066  
QY 1216 GTTCACAAAGTATTTGGCAATCTTAATCTGAGCAAGAAATAGGGGATTTT 1266  
Db 1065 GTTCACAAAGTATTTGGCAATCTTAATCTGAGCAAGAAATAGGGGATTTT 1015

RESULT 5  
LOCUS BC051053 2281 bp mRNA linear ROD 14-APR-2003  
DEFINITION Mus musculus, clone MGC:58966 IMAGE:5038182, mRNA, complete cds.  
ACCESSION BC051053  
VERSION BC051053.1 GI:29835221  
KEYWORDS MGC.  
SOURCE Mus musculus (house mouse)  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 2281)  
Strausberg, R.  
Direct Submission  
Submitted (11-APR-2003) National Institutes of Health, Mammalian  
Gene Collection (MGC), Cancer Genomics Office, National Cancer  
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,  
USA  
NIH-MGC Project URL: <http://mgc.ncl.nih.gov>  
Contact: MGC help desk  
Email: [gcgabs-remail.nih.gov](mailto:gcgabs-remail.nih.gov)  
Tissue Procurement: Gilbert Smith, Ph.D.  
CDNA Library Preparation: Life Technologies, Inc.  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNLN)  
DNA Sequencing by: National Institutes of Health Intramural  
Sequencing Center (NISC),  
Gaithersburg, Maryland;  
Web site: <http://www.nisc.nih.gov/>  
Contact: [nisc\\_mgc@nih.gov](mailto:nisc_mgc@nih.gov)  
Blakesley, N., Ayle, K., Beckstrom-Sternberg, S.M., Benjamin, B.,  
Blakesley, R.W., Boufard, G.G., Breen, K., Brinkley, C., Brooks, S.,  
Dietrich, N.L., Granite, S., Guan, X., Gupta, J., Haghighi, P.,  
Hansen, N., Ho, S.-L., Karlins, E., Kwong, P., Lalic, P., Legaspi, R.,  
Maduro, Q.L., Mastello, C., Masker, B., Mastrian, S.D., McCloskey, J.C.,  
McDowell, J., Pearson, R., Stanlipop, S., Thomas, P.J., Touchman, J.W.,  
Tsurgou, C., Vogt, J.L., Walker, M.A., Wetherby, K.D., Wiggins, L.,  
Young, A., Zhang, L.-H. and Green, E.D.

FEATURES  
Source  
Location/Qualifiers  
1..2281  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/strain="C2BEH II"  
/db\_xref="taxon:10090"  
/clone="MGC:58966 IMAGE:5038182"  
/issue\_type="Mammary tumor metastasized to lung. Tumor  
arose spontaneously from a mesodermal normal mammary  
(clonal) outgrowth infected with the virus MMTV."  
/clone\_1lb="MGI:CGAP\_Lu29"  
/lab\_host="DH10B"  
/note="Vector: pCMV-Sport6"  
197..775  
/codon\_start=1  
/product="Unknown (protein for MGC:58966)"  
/protein\_id="AAH51053.1"  
/db\_xref="GI:29835222"  
/translation="KQAKCVVVGDAVGKTCILISYTNAPPEYIPIVFDNYSNV  
MDGKPVNLGMDTAGEYDRPLPSYPTDVFLLCFSLVSPASENRYAKYPPVR  
HHCNPPIILVGTGLRDKDDEKTEKTEKLPYPOGLMAKKEIGAVKYLECSAL  
TQGLKTVGEDEAIRAVLCPPVKKRKLCIL"  
BASE COUNT 568 a 569 c 504 g 640 t  
ORIGIN

Query Match 68.6%; Score 868.2; DB 10; Length 2281;  
Best local Similarity 86.4%; Pred. No. 6.5e-218;  
Matches 1102; Conservative 0; Mismatches 123; Indels 50; Gaps 11;

QY 35 TTTTCTTTTCTTTTATGATCAAGATTTATTAAGTCATACATGCAAAACATACGCTCA 94  
Db 2281 TTTTCTTTTCTTTTATGATCAAGATTTATTAAGTCATACATGCAAAACATACGCTCA 2222  
QY 95 ACTGATTTAGCAAAAGATCAATGTAAACACATCCCAATTTGCAACTGTCAATTGAAA 154  
Db 2221 ACTGATTTAGCAAAAGATCAATGTAAACACATCCCAATTTGCAACTGTCAATTGAAA 2162  
QY 155 AAGTTTGTCTAGTGTGCGAAAGGCCCAACACTGTGTTCTGCCAGTGAATAGTTGT 214  
Db 2161 AAGTTTGTCTAGTGTGCGAAAGGCCCAACACTGTGTTCTGCCAGTGAATAGTTGT 2102  
QY 215 ACAGAACGGGTTAGCACTAGCCCTTGACACAACTCAGACACCAAGG----- 264  
Db 2101 ACAGAACGGGTTAGCACTAGCCCTTGACACAACTCAGACACCAAGG----- 2042  
QY 265 -----TACCGGAAGCATGTGTCCCGGTGGGTGAGAGGTCTAGAGGGGGCGCATCA 313  
Db 2041 GCGAGAGCCACAGACAGAACCGTCTGTCCACGTGGGAGGTCTAG-----GCGACATTA 1985  
QY 314 ATGACATGACAGTGTGTGCTGTGCGACAGAGT--GATGTTCAAGAAATATCTAAATA 371  
Db 1984 GTGACATGACAGTGTGTGCTGTGCGACAGAGT--GATGTTCAAGAAATATCTAAATA 1925  
QY 372 GTTTAAAACTGTAAAGCCGACAGCTGATTTCTACACCACTTACTAGAAAAAGCAAG 431  
Db 1924 GTTTAAAACTGTAAAGCCGACAGCTGATTTCTACACCACTTACTAGAAAAAGCAAG 1865  
QY 432 GAAGCACTAGT--CAGCTGTAAAGAGGTGAAGAGTGAAGAGCAAGCACTTCTACTATCTAC 490  
Db 1864 AAGCACTAGT--CAGCTGTAAAGAGGTGAAGAGTGAAGAGCAAGCACTTCTACTATCTAC 1805  
QY 491 C-AAAAAATCTCGAATGATGATTAAGAAAGATCTTATAGTACAGGTGAGACATATTGC 549  
Db 1804 C-AAAAAATCTCGAATGATGATTAAGAAAGATCTTATAGTACAGGTGAGACATATTGC 1745

OY	550	TCGTTAAGAAAGGGGTCCTCTAAAGAAAGACATTGCTGAATAGCAACCTGTAGAGTAGGCCC	609
Db	1744	TCGTTAAGAA- GGGGTTCTTAAGAAAGACATTGCTGAATAG- AACTGTAAAGGGTGCCC	1687
OY	610	AGTTAAATATGAGACTAACGCCCATCTGGGGAGGAGACACAGGGGGAAGGGGGGCTCA	669
Db	1686	AGTTAAATATGAGACTAACGCCCATCTGGGGAGGAGAGCGGACGAGGAGGGTGGGGGGCG	1627
OY	670	AG-----AGAGACACTATAAGATGGCCATTGTGCATCTACTGTTTGACAGA	717
Db	1626	CGAAGGATGCTCAAGAGACACTATAAGATGGCCATTGTGCATCTACTGTTTGACAGA	1567
OY	718	AATTAACCGTTAAAGAGTTTACCCTGGACCTTTATTCAGTGAATTAATCTCATGATAC	777
Db	1566	AATTAACCGTTAAAGAGTTTACCCTGGACCTTTATTCAGTGAATTAATCTCATGATAC	1507
OY	778	AATGTAGTGTAAATTAATCTCTACTCTATATATATAGTCAAAATACGTCTGTCTCCGTTGAT	837
Db	1506	AATGTAGTGTAAATTAATCTCTACTCTATATATATAGTCAAAATACGTCTGTCTCCGTTGAT	1447
OY	838	GACGTGCTGTTCAACACCTCCACCCAGACACCCAGACCTAGGAACGAATACTTCGT	897
Db	1446	CATGTGCTGTT--CACACTCCACCCAGACACCCAGACCTAGGAACGAATACTTCGAT	1389
OY	898	AGAGGCAACACAGGAGCCAGAGTCTGTCTAAACCTGACACAAAGCCGGTCAAGCTGT	957
Db	1388	AGAGGCAACACAGGAGCCAGAGTCTGTCTAAACCTGACACAAAGCTCAAGCTGTGCGT	1329
OY	958	TTAGAGAAGCTCAGTATGAATCAAGAGACAGCTGTATACACCCTC---GTGAGCTTACA	1014
Db	1328	TTAGAAAAGCTCAGTATGAATCAAGAGACAGCTGTATACACCCTCAGTGTGACGTACA	1265
OY	1015	GTACAAAGTTACGTAATGAGCATGGGCTGATTAAGTTACAGGTCGCTTACATGCAAGCGTG	1074
Db	1268	GTACAAAGTTACGTAATGAGCATGGGCTGATTAAGTTACAGGTCGCTTACATGCAACGCTG	1209
OY	1075	TCATTTAAGAGGCGTGTGTGTGTACACAGGCTGTGGAGCTACGGGAGGGTGTGCACCCCT	1134
Db	1208	TCATTTAAGAGAGCGTGTGTGTGTGTACACAGGCTGTGGAGCTATGGAAGGGGTGTGCACCCCT	1149
OY	1135	GAGCCACAGAACCTCAGTCTTCTTAAGGACAA--GTCTCTCAACAGACTTAAGTGTAC	1191
Db	1148	GAGCTCAAGAACTCAGTCTTCTTAAGCAACGAAGTCTCTCAAGAGATTAAAGTCTTAG	1089
OY	1192	GTGTTCTACACAAACGCACTTAAGTTCACAAAGGTATTTGGCAATTTCTTAATCTGACCA	1251
Db	1088	GTGTTCTACAGCAACACAACTTAAGTTCACAAAGGTATTTGGCAATTTCTTAATCTGACCA	1029
OY	1252	AGAAATAGGGGATTTT 1266	
Db	1028	AGAAATAGGGGATTTT 1014	

**TITLE**  
**JOURNAL**

**COMMENT**

```

Gordon, M., Goltz, J.S. and Kucherlapati, R.
Isoshikhe, I.P., Shih, C., Becker, J., Thomas, E., Perez, A.,
Tordella, L.,
Submitted (03-MAY-2000) Department of Molecular Genetics, Albert
Einstein College of Medicine Genome Center, 1300 Morris Park Ave.,
Bronx, NY 10461, USA
On Aug 11, 2001 this sequence version replaced gi:1493654.

-----Genome Center
Center: Harvard Partners Genome Center
Center Code: HPGC
Web site: http://www.hpgc.org/Sequence/mouse.html
Contact: hpgc@emmel.mgh.harvard.edu
-----Summary Statistics
Center project name: ABN
Sequencing vector: pUC18; L08752
Chemistry: Dye-terminator Big Dye; 100k
*Consensus quality: 256302 at least Q20
*Consensus quality: 254124 at least Q30
*Consensus quality: 250276 at least Q40
Estimated insert size: agarose-FP - N/A
**Estimated insert size: 268581 - sum-of-contigs
Quality coverage: agarose-FP - N/A
Quality coverage: 6.2 x in Q20 bases; sum-of-contigs estimation
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 26 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

```

RESULT 6	AC068493/c	269081 bp	DNA	1 linear	HTG 15-MAY-2002
LOCUS	AC068493				
DEFINITION	Mus musculus clone RP23-76K1 strain C57BL/6J, WORKING DRAFT				
ACCESSION	AC068493				
VERSION	AC068493.10	GI:15148081			
KEYWORDS	HTG; HTGS_PHASE1; HTGS_DRAFT.				
SOURCE	Mus musculus (house mouse)				
ORGANISM	Mus musculus				
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.				
AUTHORS	1 (bases 1 to 269081)				
	Montgomery,K.T., Grills,G., Han,J., Lee,E., Long,J., Pomerantz,R.,				
	Ioshikhes,I.P., Shim,C., Decker,J., Thomas,E., Pereira,A.,				
	Gordon,M., Goltz,J.S. and Kuchertlapati,R.				
TITLE	High Throughput Mouse Sequencing				
JOURNAL	Unpublished				
REFERENCE	2 (bases 1 to 269081)				
AUTHORS	Montgomery,K.T., Grills,G., Han,J., Lee,E., Long,J., Pomerantz,R.,				

*	1	51730: contig of 51730 bp in length
*	51731	51750: gap of unknown length
*	51751	82423: contig of 30673 bp in length
*	82424	82443: gap of unknown length
*	82444	118158: contig of 35715 bp in length
*	118159	118178: gap of unknown length
*	118179	139441: contig of 21263 bp in length
*	139442	139461: gap of unknown length
*	139462	160638: contig of 21177 bp in length
*	160639	160658: gap of unknown length
*	160659	177611: contig of 16553 bp in length
*	177612	177631: gap of unknown length
*	177632	190991: contig of 13360 bp in length
*	190992	191011: gap of unknown length
*	191012	200128: contig of 9117 bp in length
*	200129	200148: gap of unknown length
*	200149	211882: contig of 11734 bp in length
*	211883	211902: gap of unknown length
*	211903	220773: contig of 8871 bp in length
*	220774	220793: gap of unknown length
*	220794	229643: contig of 8850 bp in length
*	229644	229663: gap of unknown length
*	229664	238339: contig of 8676 bp in length
*	238340	238359: gap of unknown length
*	238360	244848: contig of 6489 bp in length
*	244849	244868: gap of unknown length
*	244869	249725: contig of 4857 bp in length
*	249726	249745: gap of unknown length
*	249746	252523: contig of 2778 bp in length
*	252524	252543: gap of unknown length
*	252544	257520: contig of 4977 bp in length
*	257521	257540: gap of unknown length
*	257541	260396: contig of 2856 bp in length
*	260397	260416: gap of unknown length
*	260417	262709: contig of 2293 bp in length
*	262710	262729: gap of unknown length
*	262730	263934: contig of 1205 bp in length
*	263935	263954: gap of unknown length
*	263955	264196: contig of 242 bp in length
*	264197	264216: gap of unknown length
*	264217	264629: contig of 413 bp in length

*	264530	264649:	gap of unknown length
*	264650	265608:	contig of 959 bp in length
*	265609	265628:	gap of unknown length
*	265629	266935:	contig of 1307 bp in length
*	266936	266955:	gap of unknown length
*	266956	268335:	contig of 1380 bp in length
*	268336	268355:	gap of unknown length
*	268356	268349:	contig of 194 bp in length
*	268357	268386:	gap of unknown length
*	268570	269081:	contig of 512 bp in length

**FEATURES**  
**source**

```

/organism="Mus musculus"
/mol_type="genomic DNA"
/strain="C57BL6/J"
/db_xref="taxon:10090"
/clone="RP23-76K1"
/sex="male"
1..51730
/misc_feature
  /note="assembly_name:Contig139"
51751..82423
/misc_feature
  /note="assembly_name:Contig137"
82444..118158
/misc_feature
  /note="assembly_name:Contig136"
118179..139441
/misc_feature
  /note="assembly_name:Contig135"
13462..160638
/misc_feature
  /note="assembly_name:Contig134"
160659..177611
/misc_feature
  /note="assembly_name:Contig133"
177632..190991
/misc_feature
  /note="assembly_name:Contig132"
191012..200128
/misc_feature
  /note="assembly_name:Contig131"
200149..211882
/misc_feature
  /note="assembly_name:Contig130"
211903..220773
/misc_feature
  /note="assembly_name:Contig129"
220794..229643
/misc_feature
  /note="assembly_name:Contig128"
229664..238339
/misc_feature
  /note="assembly_name:Contig127"
238360..244848
/misc_feature
  /note="assembly_name:Contig126"
244869..249725
/misc_feature
  /note="assembly_name:Contig125"
249746..252523
/misc_feature
  /note="assembly_name:Contig124
clone_end:SP6
vector_side:left"
252544..257550
/misc_feature
  /note="assembly_name:Contig123"
257541..260396
/misc_feature
  /note="assembly_name:Contig122"
260417..262709
/misc_feature
  /note="assembly_name:Contig121"
262730..263934
/misc_feature
  /note="assembly_name:Contig120"
263955..264196
/misc_feature
  /note="assembly_name:Contig119"
264217..264629
/misc_feature
  /note="assembly_name:Contig118"
264650..265608
/misc_feature
  /note="assembly_name:Contig117"
265629..266935
/misc_feature
  /note="assembly_name:Contig116"
266956..268335
/misc_feature
  /note="assembly_name:Contig115"
268356..268549
/misc_feature
  /note="assembly_name:Contig114"
268570..269081
/note="assembly_name:Contig113"
65027 a 66027 c 664672 g 67724 t 932 others

```

## ORIGIN

Query Match	68.1%;	Score 862;	DB 2;	Length 269081;
Best Local Similarity	84.8%;	Pred. No. 4,3e-216;		
Matches 1101; Conservative	0;	Mismatches 145;	Indels 52;	Gaps 10;

QY	14	TTTTCAAGTTCCAAAGCAATTTTTTTTTTTTTTTTTATGATTCACAAAGATTTATTAAGTCA	73
Db	63492	TTGGCAAAATTCCAAAAAACMACCATACACAGATGTATGATTCAAAGGATTTATTAATCA	63433
QY	74	TACATGCAAAACATACTGCTAATCGATTAGCAAAAGATCAATGTATTAACACTCCACAA	133
Db	63432	TACATGCAAAACATACTGCTAATCGATTAGCAAAAGATCAATGTATTAACACTCCACAA	63373
QY	134	TTCTGCAACTGTCATTTGAAAAAGTTTGTCTAGTGTGCAAAAGGCCACACGTGTCT	193
Db	63372	TTCTGCAACTGTCATTTGAAAAAGTTTGTCTAGTGTGCAAAAGGCCACACGTGTCT	63313
QY	194	CTTCCGCAAGTATAGTTGTACAGAAAGCCGTTTACACTAGCCGTTGACAGACCTGCAC	253
Db	63312	CTTCCGCAAGTATAGTTGTACAGAAAGCCGTTTACACTAGCCGTTGACAGACCTGCAC	63253
QY	254	AGACCCAAAG-----TACCGAAGCATGTGTCCGCTGGGTGA	292
Db	63252	AGACCCAAAGCATCTAGGACAGACCATGACAGAAAGGCTGTCTACCTAGGGGGA	63193
QY	293	GGTCTAGAGGGGGGCGCATCAATGACATGACAGTGTGTGCTACTCTGGCAACAGTGAATG	352
Db	63192	GGTCTAGA---GGGAGCATTTAGTCACATGACAGTGTGTGCTACTCTGTACAAAGTTCATAG	63136
QY	353	TTTTCAGAAATCTTAAATTAATTTAAAACTGTAAACCCGACAGCTGATTTCTACACC	412
Db	63135	TTTTCAGAAATCTTAAATTAATTTAAAACTGTAAACCCGACACACATGATTTGCACACT	63076
QY	413	AGTTACTAGAAAAAGAAAGGAGAACACTAGT-CAGCTGATTAAGAAAGGTGAAAAAGGA	471
Db	63075	AGTTGCTAGAAACTAAGAAAGCACTAAGCACTAGCTCTGAGTAAAGTAAAGCCAAAAACAGA	63016
QY	472	ACGCACCTTCACTATCTACCAAAAAATCTCCGA-----TGCAATATCAGAAAGTCT	526
Db	63015	CGCACCTTCACTAGTGCATCCAAAAAACMAAAAAATCTCTGGTATTCAGAAATCTT	62956
QY	527	ATACTACAGGTGACACATATTGCTCGTTAGAGAGGGGCTCTTAAGAAAGACACTTGCCTA	586
Db	62955	ATACTACAGGTGACACATATTGCTCGTTAGGA-GGGGTTCTTAAGAAAGACACTTGCCTA	62897
QY	587	AGTTAGCAACTGTGAGGATGGCCAGTTTAATATGCACTCMAAGCCCATCTGGGGAGGG	646
Db	62896	AGTTAG-ACCTGTGAGGATGGCCAGTTTAATATGCACTCMAAGCCCATCTGGGGAGGG	62838
QY	647	ACAGCAGAGGGGAAGGGGGGCTCAAG-----AGAGCACTGATTAAGTGGGCA	694
Db	62837	ACGGCAGTGGAGGGTGGGGGGCCGAAGGATGCTCMAAGAGACACTGATTAAGTGGGCA	62778
QY	695	TTTGTCTATCTACTGTTTGACAGAAATTAACCGTTAAAAAAGCTTTTACCCTGACACTTTTA	754
Db	62777	TTTGTCTACTACTGTTTGACAGAAATTAACCGTTAAAAAAGCTTTTACCCTGACACTTTTA	62718
QY	755	TTTCAAGTGAATTAATCATGTACAAATGTATGTATTAATCATCTCTCATATTAGTCA	814
Db	62717	TTTCAAGTGAATTAATCATGTACAAATGTATGTATTAATCATCTCTCATATTAGTCA	62658
QY	815	AAATACTGTCTGTCTCTTGTATGACGTGCTGTTTACACACTCCACCCAGCACACCAC	874
Db	62657	AAATACTGTCTGTCTCTTGTATGATCATGTCTGTTT--CACACTCCACCCAGCACACCAC	62600
QY	875	GACTAGGAAGAATTAATCGTTAGAGGCAACACAGACGCAAGTTCCTGTTCAAGCT	934
Db	62599	AACATAGGAAGAATTAATCTATTAGAGCAACACAGAAACCAAGTTCCTGTTCAAGCT	62540
QY	935	GCAGAACCGGTACGCTGATTTTATAGAGAACTCACTATGAATCAAAAGCAGAGCTGT	994
Db	62539	GCAGAACCGGTACGCTGATTTTATAGAGAACTCACTATGAATCAAAAGCAGAGCTGT	62480

Oy	995	TACACCCATC---GTGACGACGATGCAAAAGTTACGTATATGAGCATGGGCTGATAGTTA	1051
Db	62479	CACACTATATACGTGTACGTACGTCACAGTACAAAGTTACGTATATGACATGGGCTATATAGTTA	62420
Oy	1052	CAGGTGCGTACATGGCAGCGTGTCAATTAAAGAGGCTGTCTGTGCACACGCTTGGA	1111
Db	62419	CAGGTGCGTACATGGCAACGTGTCTAATTAAAGAGCGTGTGTGTGCACACGCTTGGA	62360
Oy	1112	GCTACGGGAGGGTCTGCACCCCTGACGCCACAGAAAGTCAGTCTTCTTAAGACAAA---G	1168
Db	62359	ACTCTGGAAGGGTGTGACCCCTGGCTCCAGAAAGTCGGGCTTCTTACCAACAAGAATC	62300
Oy	1169	TCTCTCAACAGCTTAGTGTCTTACGTTCTCAGCACAACGCAACTAGTCTCAAGGTAT	1228
Db	62289	TCTCAGAAGAGTTAAGTGTAGGTGTCTCAGACACAACGCAACTAGTCTCAAGAGTAT	62240
Oy	1229	TTTGGCAATTCTTATCTGAGCAAGAATAGGGATTTT	1266
Db	62239	TTTGGCAATTTTATCTGAGCAAGAATGGGGATTTT	62202

Db	1473	CTTTGATCA-----CACACACCCACCCCGGCACACCCACAGTA--AACGAATT	1426
Oy	891	CTTGCTTAGAGGCACAACAGAGCCAGATTCTGTTCAAAGCTGCAGAGCCGGTACG	950
Db	1425	CTTCATTAGAGGAATA-----GCAGTTCTGTTCAAAATCTCCGCAAAACCTGGTCAGA	1372
Oy	951	TGATATTATAGAAACTCATATGTAATCAAGAAGACAGCTGTATACCCCATGTCAG	1010
Db	1371	AACCTCGTATGMAATCATCAAAGACTGATCCAAAGAGCTGAGCTGTACGCTCACTCAT	1312
Oy	1011	TACAGTACAAAGTTACGTAATGAGCANGSGCTGATAGTACAGGTGGTTACATGGCAG	1070
Db	1311	TACAGTACAAAGTTATGTCGGGAACACGCTGCTAATCACTAGTGTGAGTTCAATGCCAA	1252
Oy	1071	CGTGCTATTAGAGGAGCGTCTGTGTGTACACGGGCTGG-----GAGCTACGGGAGGG	1123
Db	1251	CGTTCATTCGGGAGGCGTGTCTGCTTACGATCTGAGAACTACATAGAGACCAAGTGG	1192
Oy	1124	TCTGCACC-CCGTAGCCCCAAGAGCTGCAGTCTCTTTAGAGCAAAAGTCTCTCAACAGCTT	1182
Db	1191	TCTGCACCTCTTAAGTGCAGAAAGCTACGCTGTCTCCAAGACGAAGTCTTTGCCAAGTT	1132
Oy	1183	AGTGTTCAGTGTCTCTCAGCACAAC---GCAACTTAGTTCACAAAGTATTTGGCAATTC	1239
Db	1131	CAGTGCTGGGTCTTCGCGACACAATGACAGTGTAGTTTCAGAAAGTATTTTGGCAACTC	1072
Oy	1240	TTAATCTGACAGAAATAGGGG	1261
Db	1071	TTAATCTGACAAAGATGGGG	1050
RESULT 8			
HSAL32695/C			
LOCUS	HSAL32695	28567 bp	DNA
DEFINITION	Homo sapiens rac1 gene.		linear
ACCESSION	AJ132695.5		G1:8574037
VERSION	AJ132695.5		G1:8574037
KEYWORDS	alternative splicing; Alu; AluJo; Alusg1; Alusp; Alusx; AT-rich; C-rich; MIR; rac1 gene; rac1 protein; Rac1b protein; repetitive sequence.		
SOURCE	Homo sapiens (human)		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.		
AUTHORS	Matos,P., Skaug,J., Marques,B., Beck,S., Verissimo,F., Gespach,C., Boavida,M.G., Scherer,S.W. and Jordan,P.		
TITLE	Small GTPase Rac1: structure, localization, and expression of the human gene		
JOURNAL	Biochem. Biophys. Res. Commun. 277 (3), 741-751 (2000)		
MEDLINE	20517245		
PUBMED	11062023		
REFERENCE	2 (bases 1 to 28567)		
AUTHORS	Jordan,P.		
TITLE	Direct Submission		
JOURNAL	Submitted (03-FEB-1999) Jordan P., Centro de Genetica Humana, Laboratorio de Oncobiologia, Instituto Nacional de Saude "Dr. Ricardo Jorge", Avenida Padre Cruz, 1649-016 Lisboa, PORTUGAL On Jun 20, 2000 This sequence version replaced g1:7248282.		
COMMENT	Related sequence A1638561.		
FEATURES	Location/Qualifiers		
source	1..28567		
	/organism="Homo sapiens"		
	/mol_type="genomic DNA"		
	/db_xref="taxon:9606"		
	/chromosome="7"		
	/map="7p2"		
	/clone="H..NH0425A05, RPCI 11"		
	repeat_region	1..109	
	/rpt_family="Alusg"		
	repeat_region	125..422	
	/rpt_family="AlusC"		
	repeat_region	441..742	
	/rpt_family="Alusq"		
	gene	743..28117	/rpt_family="Alusq"
	/gene="rac1"	743..1226	
	promoter	911..983	
	repeat_region	936	/rpt_family="GC_rich"
	misc_feature	936	/gene="rac1"
	/note="alternative transcription start site"	1026	
	misc_feature	1037	/gene="rac1"
	/note="alternative transcription start site"	1053	
	misc_feature	1053	/gene="rac1"
	/note="alternative transcription start site"	1106..1175	
	repeat_region	1106..1175	
	/rpt_family="GC_rich"	101n(1227..1261,13701..13772,18413..18530,22804..22860	
	CDS	24268..24330,26010..26169,26458..26588)	
	/gene="rac1"		
	/note="alternative"		
	/codon_start=1		
	/product="Rac1b protein"		
	/protein_id="CAA10733.6"		
	/db_xref="GI:8574039"		
	/db_xref="SPTRMBL:095501"		
	/translation="MQAIRKCVVGDGAVGKTCLLISYTNAPGEYIPVFNDYSNV		
	LVDSKPVNIGIMDTAGGEDYDRPLSPQTVGETYKDITSGDKRPDIADVELFCS		
	VNSPASEENRAKYEPVEVRHHKCNPTPIIVGTGLDRLDDKDTLEKKELKEFTVPQ		
	GLMAKEIGAVKTELCGSLTROGLKVFEDEAIRAVICPPPVKRKRRCILL"		
	101n(1227..1261,13701..13772,18413..18530,24268..24330		
	26010..26169,26458..26588)		
	/gene="rac1"		
	/codon_start=1		
	/product="Rac1 protein"		
	/protein_id="CAB53579.5"		
	/db_xref="GI:8574038"		
	/translation="MQAIRKCVVGDGAVGKTCLLISYTNAPGEYIPVFNDYSNV		
	MVDGKPVNIGIMDTAGGEDYDRPLSPQTVGETYKDITSGDKRPDIADVELFCS		
	HCKPNPTIIVGTGLDRLDDKDTLEKKELKEFTVPQGLMAKETIGAVKTELCGSL		
	TROGLKVFEDEAIRAVICPPPVKRKRRCILL"		
	<1227..1261		
	/gene="rac1"		
	/number=1		
	1262..13700		
	/gene="rac1"		
	/number=1		
	1390..1419		
	/rpt_family="GC_rich"		
	1752..1793		
	/rpt_family="FLAM"		
	1794..2005		
	/rpt_family="AluJo"		
	2140..2171		
	/rpt_family="Simple_repeat"		
	2666..2972		
	/rpt_family="Alusx"		
	2980..3267		
	/rpt_family="Alusx"		
	3268..3289		
	/rpt_family="(T)n"		
	3802..4102		
	/rpt_family="AluB"		
	4122..4386		
	/rpt_family="Alusx"		
	4538..4719		
	/rpt_family="LTR3A"		
	4754..5053		
	/rpt_family="Alusq"		





Db	27375	TTTTTACCCTGTACACTTTCATTCCTGTGACATGTAGTGA-----AATTCACACTTGGT	27320
OY	807	A-TTAGCAAAATAACGTCTGTCTGTCTCTTGTGATACGTGCTGTTCACACATCCACCAG	865
Db	27319	ATTTTGCAAAATAACGTCTCTTGTCTCTTGTGATTA-----CACACACCCACCAG	27270
OY	866	CACACCCACGACTAGGAACAGAAATACCTGCTTAGAGGCAACACAGAGCCAGACTCTGT	925
Db	27269	CACACCCACAGCTA--AAGAAATTCCTTCAATTAGAGAAATA-----GCAGTTCGTTCC	27218
OY	926	TCAAAAGCTCTCAGAGCCGGTCACTAGCTGTGATTTTAAAGAACTCATTGAAATCAAAAG	985
Db	27217	AAAATCTCCCAAAAGGTGCTCGAAGAAACCTCGCTATGAATCACAAAAGACTGATCCAAAG	27158
OY	986	CAGAGCTGTACACACCCTCGTACAGTACAGTACAAAGTTACGTAATGAGCATGGCTGAT	1045
Db	27157	AGCGAGACTCTGACGTACCTCATCATTAACGTAACAATGTTATATGCGGGAACAGTGTCTCT	27099
OY	1046	AAGTTACAGGTGCGCTTACATGACGAGCGGTGCTATTAAAGAGCGCTGTCTGTTCACACGGT	1105
Db	27097	AAGCTACTGGTGTGATTCATGAGCAACGCTTCAATTCGGGAGGCTGTCTGTTCACCAATC	27033
OY	1106	CTGG-----GAGCTACGGGAGGAGGTGTGCAC--CGTAGCCCAAGAGCGACGTCTCT	1157
Db	27037	TGACAATCATATAGAGACAGCAAGTGTCTGCACCTCCTTAAGTCAAGAGCTACCGTCTCT	26978
OY	1158	TAAAGACAAAGTCTCTCAACAGTGTAGTGTCTTACGCTTCTCAGCAAC--GCAACTT	1214
Db	26977	CAAAAGACGAAGGCTTGTGCAAAAGTCAAGTCTCGTGAGTTCGTCGCAACAAATGCAGTGT	26918
OY	1215	AGTTCAACAAGTATTTTGGCAATTCCTTAATCTGACGACGAAGATGGGG	1261
Db	26917	AGTTCAACAAGTATTTTGGCAACTCTTAATCTGACGAAGATGGGG	26871
RESULT 9	AC009412/c		
LOCUS	AC009412	212827 bp	DNA linear PRI 07-NOV-2001
DEFINITION	Homo sapiens BAC clone RP11-425P5 from 7, complete sequence.		
ACCESSION	AC009412		
VERSION	AC009412.6	GI:14190769	
KEYWORDS	HTG.		
SOURCE	Homo sapiens (human)		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryotes; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.		
AUTHORS	1 (bases 1 to 212827)		
TITLE	Sulston, J.E. and Waterston, R.		
JOURNAL	Toward a complete human genome sequence		
REFERENCE	Genome Res. 8 (11), 1097-1108 (1998)		
AUTHORS	99063792		
TITLE	9847074		
JOURNAL	2 (bases 1 to 212827)		
REFERENCE	Hou, S., Maupin, R., Haekenson, W., Gregory, S. and Belter, E.		
AUTHORS	The sequence of Homo sapiens BAC clone RP11-425P5		
TITLE	Unpublished		
JOURNAL	3 (bases 1 to 212827)		
REFERENCE	Waterston, R.H.		
AUTHORS	Direct Submission		
TITLE	Submitted (21-APR-1999) Genome Sequencing Center, Washington		
JOURNAL	University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA		
REFERENCE	4 (bases 1 to 212827)		
AUTHORS	Waterston, R.H.		
TITLE	Direct Submission		
JOURNAL	Submitted (23-MAY-2001) Genome Sequencing Center, Washington		
REFERENCE	University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA		
AUTHORS	5 (bases 1 to 212827)		
TITLE	Waterston, R.		
JOURNAL	Direct Submission		
REFERENCE	Submitted (07-NOV-2001) Department of Genetics, Washington		
AUTHORS	University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA		

```

COMMENT
-----
On May 23, 2001 this sequence version replaced g1:13431187.

----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc
Contact: saplens@wustl.wustl.edu
----- Summary Statistics -----
Center project name: H_NN0425P05

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:
all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:
The sequence of this clone was established as part of a mapping and sequencing collaboration between the NHGRI Chromosome 7 Mapping Project (Eric D. Green, Director), John D. McPherson in the Department of Genetics (Washington University), and the Washington University Genome Sequencing Center. For additional information about the map position of this sequence, see
http://www.nhgri.nih.gov/DIR/CTB/CHR7 , send
mailto:egreen@nhgri.nih.gov , or see http://genome.wustl.edu/gsc

SOURCE INFORMATION:
The RPCI-11 human BAC library was made from the blood of one male donor, as described by Osoegawa,K., Woon,P.Y., Zhao,B., Frengen,E., Tateo,M., Catanese,J.J. and de Jong,P.U. (1998) An Improved Approach for construction of bacterial artificial chromosome libraries. Genomics 51:1-8. The clone may be obtained either from Research Genetics, Inc. (http://www.resgen.com) or Pieter de Jong and coworkers at the Roswell Park Cancer Institute (http://daccpac.med.buffalo.edu)
VECTOR: pRACE3.6

NEIGHBORING SEQUENCE INFORMATION:
The clone sequenced to the right is CTD-2195F2, 200 bp overlapi. Actual start of this clone is at base position 1 of RP11-425P5; actual end is at base position 212633 of RP11-425P5.

Location/Qualifiers
1..212827
    /organism="Homo sapiens"
    /mol_type="genomic DNA"
    /db_xref="taxon:9606"
    /chromosome="7"
    /map="7"
    /clone="RP11-425P5"
    /clone_1lb="RPCI-11"
    1623..1696
        /rpt_family="MIR"
    1698..2006
        /rtp_family="Alu"
    2024..2337
        /rtp_family="Alu"
    3006..3093
        /rtp_family="L1"
    3453..3552
        /rtp_family="L2"
    4924..5046
        /rtp_family="ERV.L"
    5318..5693
        /note="similar to EST BF996284 (NID:g12402607)"
    5580..5845
        /rtp_family="Alu"
    5925..6277
repeat_region

```



Db 175463 TGAATATACCAACCTTTCTTTCATCATCTGACATTGACAGAAATTAACCTTTAGAG 175404  
OY 750 TTTTATTCAGTTGAA---TTACTCCATGTACATAGTAGTAATTAATCTACTTCAAT 806  
Db 175403 TTTTATACCGGTGACACTTTCATTCCTGTGACATATAGTGA-----AATCTCCACTTCGT 175348  
OY 807 A-TTATGCAAAATCTGCTGTCTGCTCTGTTGANGACGTCTGTTTCAACACATCCACCGAG 865  
Db 175347 ATTCTGTCAAAATACGTCTGTCTGCTTGTGATCA-----CACACACCCACCCCG 175298  
OY 866 CACACCCAGACAGAGACAGAAATCTGTTAGAGCAACACAGAGACGAGTCTGT 925  
Db 175297 CACACCCACAGCTA--AACAGATTTCTTCAATGAGAAATA-----GCAGTCTGTTC 175246  
OY 926 TCAAAACCTGCAGAACCCGCTGCTGCTATTTTACAGAACTCATATGAAATCAAAAGAG 985  
Db 175245 AAAATCTCCGCAAAACCTGTGTGAGAAACCTCGCTATGAAATCAAAAGACTGATCCAAAG 175186  
OY 986 CAGAGCTGTACACCATCTGTGACGTACAGTACAAAGTACATAGACATGAGCTGAT 1045  
Db 175185 AGCTGAGCTGTACAGCTCACTCCATTAAGTACAAAGTATGTCGGGAACACGTGCTGT 175126  
OY 1046 AAGTTACAGTGTGCTTACATGACGAGCTGTCTATTAAGAGAGCTGTGCTGTACACAGT 1105  
Db 175125 AACTCAGTGTGATGATCATATGCGAAGCTTCAATGCGAGAGCTGTCTGTTCACGATC 175066  
OY 1106 CTGG-----GAGTACGAGGAGGCTGTGCACC-CTGAGCCACAGAGCTGACGTCTTC 1157  
Db 175065 TGAGAACTACATAGGAGACCAAGTGTGCTGCTGCTTAACTGACAGAGCTGCTGTTC 175006  
OY 1158 TTAGGACAAAGTCTCAACAGCTGTGCTGTGCTGTCTGACCAAC--GCAACT 1214  
Db 175005 CAAGACGAGAGGCTTTTGCAGAACTTCAAGTGTGCTGCTGTCTGCTGCTGCTGCT 174946  
OY 1215 AGTTCAAGAGATTTTGGCAATTCCTTAATCTGAGCAAGAAATAGGG 1261  
Db 174945 AGTTCAAGAGATTTTGGCACTTAATCTGAGCAAGAAATAGGG 174899

RESULT 10  
AK054993/c 2315 bp mRNA linear PRI 01-JAN-2002  
LOCUS Homo sapiens cDNA FLJ30431 fis, clone BRACE2008968, highly similar  
DEFINITION to RAS-RELATED C3 BOTULINUM TOXIN SUBSTRATE 1.  
ACCESSION AK054993  
VERSION AK054993.1 GI:16549633  
KEYWORDS oligo capping; fis (full insert sequence).  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE  
AUTHORS 1 Tashiro, H., Yamazaki, M., Watanabe, K., Kumagai, A., Itakura, S.,  
Fukuzumi, Y., Fujimoto, Y., Komiyama, M., Sugiyama, T., Irie, R.,  
Otsuki, T., Sato, H., Wakamatsu, A., Ishii, S., Yamamoto, J., Isono, Y.,  
Kawai-Hiro, Y., Saito, K., Nishikawa, T., Kimura, K., Yamashita, H.,  
Matsumoto, K., Nakamura, Y., Sekine, M., Kikuchi, H., Kanda, K.,  
Wagatsuma, H., Murakawa, K., Kanemori, K., Takahashi-Fujii, A.,  
Oshima, A., Sugiyama, A., Kawakami, B., Suzuki, Y., Sugano, S.,  
Nagahara, K., Masuno, Y., Nagai, K., and Isogai, T.  
NEBO human cDNA sequencing project  
Unpublished  
2 (bases 1 to 2315)  
TITLE Isogai, T., Otsuki, T. and Sugiyama, T.  
REFERENCE Direct Submission  
AUTHORS Submitted (24-OCT-2001) Takao Isogai, Helix Research Institute,  
Genomics Laboratory; 1532-3 Yana, Kisarazu, Chiba 292-0812, Japan  
(E-mail: genomics@hri.co.jp, Tel:81-438-52-3975, Fax:81-438-52-3966)  
COMMENT NEBO human cDNA sequencing project supported by Ministry of  
Economy, Trade and Industry of Japan; cDNA full insert sequencing:  
Research Association for Biotechnology (RAB) (supported by Japan  
Construction: Helix Research Institute (HRI) (supported by Japan  
Key Technology Center etc.); 5'- & 3'-end one pass sequencing: RAB,

HRI, and Biotechnology Center, National Institute of Technology and  
Evaluation; clone selection for full insert sequencing: RAB and  
HRI.

FEATURES  
source Location/Qualifiers  
1..2315  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="BRACE2008968"  
/issue\_type="cerebellum"  
/clone\_lib="BRACE2"  
/note="cloning vector: pME185FL3"  
BASE COUNT 603 a 531 c 501 g 680 t  
ORIGIN

Query Match 34.0%; Score 430.6; DB 9; Length 2315;  
Best Local Similarity 67.8%; Pred. No. 1.8e-102;  
Matches 859; Conservative 0; Mismatches 324; Indels 84; Gaps 15;

OY 50 ATGATTCAGAGATTTATTAAGTCATATGCAAAACATCTGCTAATGCAATTAAGCAAA 109  
Db 2315 ATGATTCAGAGATTTATTAAGTCATATGCAAAACATTAATGCTAATGCAATTAAGCAAA 2256  
OY 110 GATCAATGTAAACACCTCCAAATCTGCACTGCTCAATGAAAAGTTGTCTACT 169  
Db 2255 GATCAATGTAAACACCTCCAAATCTGCACTGCTCAATGAAAAGTTGTCTACT 2196  
OY 170 GGTGAAAGGCCCAACACTGTGTTCTGCGAGTGTAGTGTGTGACAGAACGGCTTAC 229  
Db 2195 GGTGAAAGGCCCAACACTGTGTTCTGCGAGTGTAGTGTGTGACAGAACGGCTTAC 2136  
OY 230 CACTACGCG--TTGACAGAACTCAGACAGCCAAAG-----TA 266  
Db 2135 CACTACAGAGTTTAAAGAACCTCAGACAGCCAAAG-----TA 2076  
OY 267 CCGAAGACATGTGCGCGGGGTGAGGCTAGAGGGGGGCGATCAATCAATGACAT 326  
Db 2075 CCGAAGACATGTGCGCGGGGTGAGGCTAGAGGGGGGCGATCAATCAATGACAT 2017  
OY 327 GTTGTACTCTGCGCAAGACAGTATG-----TTTCAGATATCTTAA 369  
Db 2016 GTTGTACTCTGCGCAAGACAGTATG-----TTTCAGATATCTTAA 1957  
OY 370 TATTTTAAACCTGTAAAGCGCGACAGCTGATTTCTACACCTGATTAAGAAACGA 429  
Db 1956 TATTTTAAACCTGTAAAGCGCGACAGCTGATTTCTACACCTGATTAAGAAACGA 1897  
OY 430 GGGAGCAGTACGTACAGTGTAAAGGAGTAAAGAGGAGAAACAGAGCAGCTCTACTATCA 489  
Db 1896 GGGAGCAGTACGTACAGTGTAAAGGAGTAAAGAGGAGAAACAGAGCAGCTCTACTATCA 1842  
OY 490 CCAAAAAATCTCCGATGCAATTCAGAAAGATCTTATAGTACAGCTCAGACATATTCG 549  
Db 1841 CCAAAAAATCTCCGATGCAATTCAGAAAGATCTTATAGTACAGCTCAGACATATTCG 1783  
OY 550 TCGTTTAAAGAGGGGCTCTAAAGAAAGCACTGCTTAAGTACCACTGTGAGATGGC 609  
Db 1782 TCGTTTAAAGAGGGGCTCTAAAGAAAGCACTGCTTAAGTACCACTGTGAGATGGC 1723  
OY 610 AGTTTAAATGTGAGCTCAACGCCCATCTGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 669  
Db 1722 AGTTTAAATGTGAGCTCAACGCCCATCTGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1663  
OY 670 AGAGAGACACTGTAAAGATCGCCATTTGTCTACTGTTGACAGAAATTAACCGTTA 729  
Db 1662 AGAGAGACACTGTAAAGATCGCCATTTGTCTACTGTTGACAGAAATTAACCGTTA 1603  
OY 730 AAAAGCTTAAACCGTGACACTTTTATTAAGTTGAA---TTACTCAGTACATATGAGG 786  
Db 1602 CAGAAATTAACCTTTTAAAGATTTTACCCTGACCTTTCATTCCTTGAATTAAGG 1543  
OY 787 TAAATTAATCTACTTCAATA-TTAAAGCAAAATAGTGTGCTGCTTGAATGAGCTGT 845





pieces.  
AC102775.3 GI:29164654  
HTG: HTGS PHASE1; HTGS DRAFT; HTGS-FULLTOP.  
KEYWORDS  
SOURCE  
Mus musculus (house mouse)  
ORGANISM  
Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
REFERENCE  
AUTHORS  
1 (bases 1 to 185245)  
Birren, B., Nusbaum, C. and Lander, E.  
TITLE  
Mus musculus, clone RP23-115C10  
JOURNAL  
REFERENCE  
2 (bases 1 to 185245)  
Birren, B., Linton, L., Nusbaum, C., Lander, E., All, A., Allen, N.,  
Anderson, S., Barna, N., Bastien, V., Boguslavsky, L., Boukhalter, B.,  
Brown, A., Camarata, J., Campopiano, A., Chang, J., Chazaro, B.,  
Choepel, Y., Colangelo, M., Collins, S., Collymore, A., Cook, A.,  
Cooke, P., Dattellano, K., Dewar, K., Diaz, J.S., Dodge, S., Faro, S.,  
Ferreira, P., Fitzhugh, W., Gage, D., Galagan, J., Gaidyna, S.,  
Ginde, S., Gord, S., Goyette, M., Graham, L., Grand-Pierre, N.,  
Hagos, B., Heatford, A., Horton, L., Hulme, W., Iliev, I., Johnson, R.,  
Jones, C., Kamat, A., Karatas, A., Kells, C., LaRocque, K.,  
Lamasares, R., Landers, T., Lehoczy, J., Levine, R., Liu, G.,  
Maclean, C., MacDonald, P., Major, J., Marquis, N., Matthews, C.,  
McCarthy, M., McEwan, P., McKernan, K., McPheeters, R., Meldrum, J.,  
Meneus, L., Mihova, T., Mienga, V., Murphy, T., Naylor, J., Nguyen, C.,  
Norbu, C., Norman, C.H., O'Connor, T., O'Donnell, P., O'Neill, D.,  
O'Neil, J., Peterson, K., Phukhang, P., Pierre, N., Pollara, V.,  
Raymond, C., Retta, R., Rieback, M., Riley, R., Rise, C., Rogov, P.,  
Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, S., Schupbach, R.,  
Seaman, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N.,  
Strauss, N., Subramanian, A., Talamas, J., Testafaye, S., Theodore, J.,  
Topham, K., Travers, M., Travis, N., Triggillo, J., Vassiliev, H.,  
Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J., Young, G.,  
Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.  
TITLE  
Direct Submission  
JOURNAL  
Submitted (23-NOV-2001) Whitehead Institute/MIT Center for Genome  
REFERENCE  
Research, 320 Charles Street, Cambridge, MA 02141, USA  
AUTHORS  
3 (bases 1 to 185245)  
Birren, B., Nusbaum, C., Lander, E., Abouelell, A., Allen, N.,  
Anderson, S., Arachchi, H.M., Barna, N., Bastien, V., Bloom, T.,  
Boguslavsky, L., Boukhalter, B., Camarata, J., Chang, J., Choepel, Y.,  
Collymore, A., Cook, A., Cooke, P., Corum, B., Dattellano, K.,  
Diaz, J.S., Dodge, S., Dooley, K., Dorris, L., Erickson, J., Faro, S.,  
Ferreira, P., Fitzgerald, M., Gage, D., Galagan, J., Gaidyna, S.,  
Graham, L., Grand-Pierre, N., Hafez, N., Hagopian, D., Hagos, B.,  
Hall, J., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C.,  
Kamat, A., Karatas, A., Kells, C., Landers, T., Levine, R.,  
Lindblad-Toh, K., Liu, G., Lui, A., Mabbitt, R., Maclean, C.,  
MacDonald, P., Major, J., Manning, J., Matthews, C., McCarthy, M.,  
Meldrum, J., Meneus, L., Mihova, T., Mienga, V., Murphy, T., Naylor, J.,  
Nguyen, C., Nicol, R., Norbu, C., O'Connor, T., O'Donnell, P.,  
O'Neill, D., Oliver, J., Peterson, K., Phukhang, P., Pierre, N.,  
Rachupa, A., Ramsamy, V., Raymond, C., Retta, R., Rise, C., Rogov, P.,  
Roman, J., Schauer, S., Schupbach, R., Seaman, S., Severy, P., Smith, C.,  
Spencer, B., Stange-Thomann, N., Stojanovic, N., Stubbs, M.,  
Talamas, J., Testafaye, S., Theodore, J., Topham, K., Travers, M.,  
Vassiliev, H., Venkataraman, V.S., Viel, R., Vo, A., Wilson, B., Wu, X.,  
Wyman, D., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.  
TITLE  
Direct Submission  
JOURNAL  
Submitted (23-MAR-2003) Whitehead Institute/MIT Center for Genome  
COMMENT  
Research, 320 Charles Street, Cambridge, MA 02141, USA  
On Mar 23, 2003 this sequence version replaced gi:22381790.  
All repeats were identified using RepeatMasker:  
Smit, A.F.A. & Green, P. (1996-1997)  
http://ftp.genome.washington.edu/RW/RepeatMasker.html  
----- Genome Center  
Center: Whitehead Institute/ MIT Center for Genome Research  
Center code: WIRB  
Web site: http://www-seq.wi.mit.edu  
Contact: sequence\_submissions@genome.wi.mit.edu  
----- Project Information  
Center project name: I19647

Center clone name: 115\_C\_10  
----- Summary Statistics  
Sequencing vector: Plasmid, n/a; 100% of reads  
Chemistry: Dye-terminator Big Dye; 100% of reads  
Assembly program: Phrap; version 0.960731  
Consensus quality: 183758 bases at least Q40  
Consensus quality: 184136 bases at least Q30  
Consensus quality: 184291 bases at least Q20  
Insert size: 178000; agarose-fp  
Insert size: 184445; sum-of-ctgts  
Quality coverage: 12.6 in Q20 bases; agarose-fp  
Quality coverage: 12.2 in Q20 bases; sum-of-ctgts  
-----  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 9 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.  
\*  
\* 1  
\* 53693: contig of 53693 bp in length  
\* 53793: gap of 100 bp  
\* 53794: 56675: contig of 2862 bp in length  
\* 56676: 56775: gap of 100 bp  
\* 56776: 60589: contig of 3814 bp in length  
\* 60590: 60690: gap of 100 bp  
\* 60690: 67210: contig of 6521 bp in length  
\* 67211: 67310: gap of 100 bp  
\* 67311: 82585: contig of 15275 bp in length  
\* 82586: 82685: gap of 100 bp  
\* 82686: 103118: contig of 20433 bp in length  
\* 103119: 103218: gap of 100 bp  
\* 103219: 132145: contig of 28227 bp in length  
\* 132146: 132245: gap of 100 bp  
\* 132246: 182335: contig of 50090 bp in length  
\* 182336: 182435: gap of 100 bp  
\* 182436: 185245: contig of 2810 bp in length.  
\*  
\* Location/Qualifiers  
\* 1. 185245  
\* /organism="Mus musculus"  
\* /mol\_type="genomic DNA"  
\* /db\_xref="taxon:10090"  
\* /clone="RP23-115C10"  
\* /clone\_1lb="RPCT-23 Female Mouse BMC"  
\* 1. 53693  
\* /note="assembly-fragment  
\* clone\_end:SP6  
\* vector\_side:left"  
\* 53794. .56675  
\* /note="assembly-fragment"  
\* 56776. 60589  
\* /note="assembly-fragment"  
\* 60690. .67210  
\* /note="assembly-fragment"  
\* 67311. .82585  
\* /note="assembly-fragment"  
\* 82686. 103118  
\* /note="assembly-fragment"  
\* 103219. .132145  
\* /note="assembly-fragment"  
\* 132246. .182335  
\* /note="assembly-fragment"  
\* 182436. .185245  
\* /note="assembly-fragment  
\* clone\_end:T7  
\* vector\_side:right"  
BASE COUNT 56870 a 34956 c 35069 g 57550 t 800 others  
ORIGIN  
Query Match 32.2% Score 408.2; DB 2; Length 185245;  
Best Local Similarity 89.8% Pred. No. 2.2e-96;  
Matches 486; Conservative 0; Mismatches 43; Indels 12; Gaps 4;

QY	736	TTTACCCGGTACGACCTTTTAAATTCAGTTGAATTAACACATGACATGAGTGAATTAAT	795
Db	607	TTCCCCCATGACACTTTTATTATTCAGTTGAATTAACACATGAGTGAATTAAT	548
QY	796	CTCATCTTTCATATTAGTCAAAATA----CTGTCTGTCTCTTTTGATGACGTCTGTATCA	851
Db	547	CTCATCTTTCATATTAGTCAAAATAACTGTCTGTCTCTTTTGATGACGTCTGTATCA	490
QY	852	CACACTCCACCACACAGACACCCACAGACTTGAAGAAAGAAATCTTCTTTAGAGCAACACAG	911
Db	489	CACACTCCACCACACACACCCACAGAACTGAAGAAAGAAATCTTTCATGAGAGCAACACACG	430
QY	912	AGCCAGAGTTCTGTTCAAAAGCCTCAGAAAGCCGCTCAGCTGTATTTAGAGAACTACT	971
Db	429	AACCAAGATCTGTTCAAAAGTCTCMAAAGCTACTGCGCTGTCTTTAGAAAACCTACT	370
QY	972	ATGAATTCMAAGACAGCAGCTGTTCACACCCATC--GTGACGTACAGTACAAATTAAGT	1028
Db	369	ATGAATTCMAAGACAGCTGTGCTACACTCTACTGTGACGTACAGTATTAAGTAAATTAAGT	310
QY	1029	AATAGCATGGGCGATTAAGTTAAGAGTGGCTTACATGGCAGCGCTGTATTAAGAGGCT	1088
Db	309	AATAGCATGGGCGATTAAGTTAAGAGTGGCTTACATGGCAGCGCTGTATTAAGAGGCT	250
QY	1089	GTGCTGTGTACACAGGCTGTGGAGCTACGAGAGGCTGCACCCCTGACCCCAAGAGCT	1148
Db	249	GTGCTGTGTACACAGGCTGTGGAGGACTATGAAAGGCTGTGCACCCCTGACTCTCAAAAGCT	190
QY	1149	CAGTCTTCTTAAGACAAA--GTCTCTCAACAGCTTAGTCTTAAGTCTTCTCAGACACA	1205
Db	189	CAGTCTTCTTAAAGACAAAAGTCTCTCAACAGAGTTAGTCTTAAGTCTTCTCAGACACA	130
QY	1206	ACGCACTTACTTCAACAAGTATTTGGCAATCTTAACTGACACAAAGAGGATTT	1285
Db	129	ACGCACTTACTTCAACAAGTATTTGGCAATCTTAACTGACACAAAGAGGATTT	70
QY	1266	T 1266	
Db	69	T 69	

RESULT 14  
AL583884/c  
LOCUS  
DEFINITION  
Accession  
Version  
Keywords  
Source  
Organism  
Reference  
Authors  
Title  
Journal  
Comment

AL583884 202342 bp DNA linear ROD 24-DEC-2002  
Mouse DNA sequence from clone RP23-324B16 on chromosome 15,  
complete sequence.  
AL583884  
AL583884.20 GI:27368255  
HTG.  
Mus musculus (house mouse)  
Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 202342)  
Smith,M.  
Direct Submission  
Submitted (21-DEC-2002) Wellcome Trust Sanger Institute, Hinxton,  
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:  
humquerry@sanger.ac.uk  
On Dec 23, 2002, this sequence version replaced gi:26788018.  
Sequence from the Mouse Genome Sequencing Consortium whole genome  
shotgun may have been used to confirm this sequence. Sequence data  
from the whole genome shotgun alone has only been used where it has  
a phred quality of at least 30.

----- Genome Center  
Center: Wellcome Trust Sanger Institute  
Center code: SC  
Web site: http://www.sanger.ac.uk  
Contact: humquerry@sanger.ac.uk  
-----

During sequence assembly data is compared from overlapping clones.  
Where differences are found these are annotated as variations

[illegible]



DB 184513 T 184513

RESULT 15  
AC105979

LOCUS Mus musculus clone RP24-301E12, WORKING DRAFT SEQUENCE, 21 ordered pieces.

DEFINITION AC105979 192498 bp DNA linear HTG 21-OCT-2002

AC105979

AC105979.4 GI:24182089

HTG: HTGS\_PHASE2: HTGS\_DRAFT; HTGS\_FULLTOP.

KEYWORDS Mus musculus (house mouse)

SOURCE Mus musculus

ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 192498)

AUTHORS Birren, B., Nusbaum, C. and Lander, E.

TITLE Mus musculus, clone RP24-301E12

JOURNAL Unpublished

AUTHORS 2 (bases 1 to 192498)

Birren, B., Linton, L., Nusbaum, C., Lander, E., All, A., Allen, N., Anderson, S., Barna, N., Bastien, V., Boguslavskiy, L., Boukhgalter, B., Brown, A., Camarata, J., Campopiano, A., Chang, J., Chazaro, B., Choepel, Y., Colangelo, M., Collins, S., Collymore, A., Cook, A., Cooke, P., Dearellano, K., Dewar, K., Diaz, J. S., Dodge, S., Ferreira, P., FitzHugh, W., Gage, D., Galagan, J., Gardyna, S., Glade, S., Gord, S., Goyette, M., Graham, L., Grand-Pierre, N., Hagos, B., Heatford, A., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Kamal, A., Karatas, A., Kells, C., Larocque, K., Lamazares, R., Landers, T., Lehoczy, J., Levine, R., Liu, G., Maclean, C., Macdonald, P., Major, J., Marquis, N., Matthews, C., McCarthy, M., McKeown, P., McKernan, K., McPheeters, R., Meldrum, J., Menus, L., Mihova, T., Mienga, V., Murphy, T., Naylor, J., Nguyen, C., Norbu, C., Norman, C. H., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K., Phunkhang, P., Plerre, N., Pollara, V., Raymond, C., Retta, R., Rieback, M., Riley, R., Rise, C., Rogov, P., Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, S., Schupback, R., Seaman, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N., Strauss, K., Subramanian, A., Talamas, J., Testaye, S., Theodore, J., Topham, K., Travers, M., Travis, N., Triggillo, J., Vassiliev, G., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W. D., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

TITLE Direct Submission

JOURNAL Submitted (11-JAN-2002) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA

REFERENCE 3 (bases 1 to 192498)

AUTHORS Birren, B., Nusbaum, C., Lander, E., All, A., Allen, N., Anderson, S., Barna, N., Bastien, V., Bloom, T., Boguslavskiy, L., Boukhgalter, B., Camarata, J., Chang, J., Chazaro, B., Choepel, Y., Collymore, A., Cook, A., Cooke, P., Dearellano, K., Dewar, K., Diaz, J. S., Dodge, S., Faro, S., Ferreira, P., FitzGerald, M., Gage, D., Galagan, J., Gardyna, S., Gord, S., Graham, L., Grand-Pierre, N., Hagos, B., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Kamal, A., Karatas, A., Kells, C., Landers, T., Levine, R., Lindblad-Toh, K., Liu, G., Maclean, C., Macdonald, P., Major, J., Matthews, C., McCarthy, M., Meldrum, J., Menus, L., Mihova, T., Mienga, V., Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Norbu, C., Norman, C. H., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K., Phunkhang, P., Plerre, N., Raymond, C., Retta, R., Rise, C., Rogov, P., Roman, J., Roy, A., Schauer, S., Schupback, R., Seaman, S., Severy, P., Smith, C., Spencer, B., Stange-Thomann, N., Stojanovic, N., Talamas, J., Testaye, S., Theodore, J., Topham, K., Travers, M., Vassiliev, G., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

TITLE Direct Submission

JOURNAL Submitted (21-OCT-2002) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA

COMMENT On Oct 21, 2002 this sequence version replaced g1:2019424. All repeats were identified using RepeatMasker: Smit, A.F.A. & Green, P. (1996-1997) <http://ftp.genome.washington.edu/RM/RepeatMasker.html>

Center: Whitehead Institute/ MIT Center for Genome Research  
Center code: MIBR  
Web site: <http://www-seq.wi.mit.edu>  
Contact: [sequence\\_submissions@genome.wi.mit.edu](mailto:sequence_submissions@genome.wi.mit.edu)

----- Project Information -----  
Center project name: L20199  
Center clone name: 301\_E.12

----- Summary Statistics -----  
Sequencing vector: Plasmid; n/a; 100% of reads  
Chemistry: Dye-terminator Big Dye; 100% of reads  
Assembly program: Phrap; version 0.960731  
Consensus quality: 190070 bases at least Q40  
Consensus quality: 191243 bases at least Q30  
Consensus quality: 191514 bases at least Q20  
Insert size: 163000; agarose-fp  
Insert size: 191698; sum-of-contigs  
Quality coverage: 11.6 in Q20 bases; sum-of-contigs  
Quality coverage: 9.9 in Q20 bases; sum-of-contigs

\*\*\*\*\*  
NOTE: This is a 'working draft' sequence. It currently consists of 9 contigs. Gaps between the contigs are represented as runs of N. The order of the pieces is believed to be correct as given, however the sizes of the gaps between them are based on estimates that have been provided by the submitter.  
\* This sequence will be replaced  
\* by the finished sequence as soon as it is available and the accession number will be preserved.

1 1061: contig of 1061 bp in length  
1 1062: gap of 100 bp  
\* 1162: contig of 4553 bp in length  
\* 5715: gap of 100 bp  
\* 5815: contig of 5608 bp in length  
\* 11423: gap of 100 bp  
\* 11523: contig of 6078 bp in length  
\* 17601: gap of 100 bp  
\* 17701: contig of 12457 bp in length  
\* 30158: gap of 100 bp  
\* 30258: contig of 8550 bp in length  
\* 38788: gap of 100 bp  
\* 38888: contig of 60357 bp in length  
\* 99445: gap of 100 bp  
\* 99445: contig of 43719 bp in length  
\* 143264: gap of 100 bp  
\* 143264: contig of 49135 bp in length.  
\* 143364: 192498: contig of 49135 bp in length.

FEATURES  
source  
1..192498  
/organism="Mus musculus"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:10090"  
/clone="RP24-301E12"  
/clone\_1lb="RP24-301E12"  
1..1061  
/note="assembly-fragment"  
misc\_feature  
1162..5714  
/note="assembly-fragment"  
misc\_feature  
5815..11422  
/note="assembly-fragment"  
misc\_feature  
11523..17600  
/note="assembly-fragment"  
misc\_feature  
17701..30157  
/note="assembly-fragment"  
misc\_feature  
30258..38787  
/note="assembly-fragment"  
misc\_feature  
38888..99444  
/note="assembly-fragment"  
misc\_feature  
99545..143263  
/note="assembly-fragment"  
misc\_feature  
143364..192498  
/note="assembly-fragment"  
clone\_end: 17  
vector\_side:right

BASE COUNT 53596 a 42471 c 43252 g 52379 t 800 others

**ORIGIN**

Query Match	29.2%	Score 369.4	DB 2:	Length 192498;
Best Local Similarity	83.6%	Pred. No. 3.9e-86;		
Matches 494;	Conservative 0;	Mismatches 56;	Indels 41;	Gaps 5;

OY	35	TTTTTTTTTTTTTTTTATGATTCACAGATTTATATTAATCTATACCTGCAAAACATACGCTA	94
Db	90240	TTTTTTTTTTTTTAAATGATTCACAGATTTATTAATCTATACCTGCAAAACATACGCTA	902399
OY	95	ACTGCATTACCAAAAGATCAATGTAAAAACATCCACAAATTCGCAACTGTCAATTGANA	154
Db	90300	ACTGCATTACCAAAAGATCAATGTAAAAACATCCACAAATTCGCAACTGTCAACTGANA	903599
OY	155	AAAGTTGTTCTAGTGTGCAAAAGGCCAACACTGTGTTCTTGCCAGTGAAGTTAGTTCT	214
Db	90360	AAAGTTGTTCTAGTGTGTTAAAGGCCAACACTGTATTCTTGCCAGTGAAGTTAGTTCT	90419
OY	215	ACAGAACGGGGTAAAGCACTAGCGCTGACGAACCTCACAGACCACCAAGG-----	264
Db	90420	ACAGAACCAATTAAGCACTAGCACTAGCAAGAACTCACAGACCCCAAGAACATCTCTA	90479
OY	265	-----TACCGAAGCATGTGTCCGCTGGGTGAGAGTCTAAGAGGGGGCGCATCA	313
Db	90480	GGCAGAGCCACGACAGAGAACGTCGTGCTGCGTAGAGAGCTTAA - GAGGAGCATTA	90538
OY	314	ATCCATGACAGTGTGTGTACTCTGGCAGACAGTATGTTTATAGAAATCTTAAATACT	373
Db	90539	GTCCATGACAGTGTGTGTAAGAGTTCTA-----GTTTAAAGTAATCTTAAATACT	90589
OY	374	TTAAAAACGTAAAGCCGACAGCAAGTATTTTCACACCCAGTTACTATAGAAACGAAGGA	433
Db	90590	TTAAAAACGTAAAGCTGCAACACATGATTTGCAACACCTTAGTGTCTAGAAACTTAAGGA	90649
OY	434	AGCACTAGTCAGC--TGAGTAAAGGAAGTGAAACAGAACGCACCTTACT-----	484
Db	90650	AGCACTAAGTACGTTTGATTAAGTAAAGTAAAGTAAACAGAACGCCGTTCTACGATCGAC	90709
OY	485	-ATTACCAAAAAAATTCGGAATGCATTTCAGAAAGATCTTATAGTACAGGTGACACA	543
Db	90710	CAAAAAAATAAATAATCTCTGAATGCAATTATCAGAAAGATCTTATAGTACAGGTGACACA	90769
OY	544	TATTTGCTGTTAAGAAAGGGGGTCTTAAAGAAAGCACTTGCTTAAGTTAGCA	594
Db	90770	TATTTGCTATTAAAGAGGGGGTCTTAAAGAAAGCACTTGCTTAAGTTAGCA	90820

Search completed: August 26, 2003, 02:29:57  
Job time : 4683.87 secs





XX WP1: 2001-355948/37.  
DR  
XX  
PT Screening hepatotoxic agent comprises contacting test cell population  
PT expressing RISKMARKER or INJURYMARKER with agent, comparing expression  
PT with reference population and identifying difference in expression  
PT levels  
XX  
PS Disclosure: Page 7-8; 76pp; English.  
XX  
CC The present invention describes a method of screening a test agent for  
CC hepatotoxicity. The method comprises: (a) providing a test cell  
CC population comprising a cell capable of expressing one or more nucleic  
CC acid sequences selected from the group consisting of RISKMARKER 1-8  
CC and INJURYMARKER 1-10; (b) contacting the test cell population with a  
CC test agent; (c) measuring expression of one or more of the nucleic  
CC acid sequences in the test cell population; (d) comparing the  
CC expression of the nucleic acid sequence in the test cell population to  
CC the expression of the nucleic acid sequence in an reference cell  
CC population comprising at least one cell whose exposure status to a  
CC hepatotoxic agent is known; and (e) identifying a difference in  
CC expression levels of the RISKMARKER or INJURYMARKER sequences, if  
CC present, in the test cell population and reference cell population.  
CC The method is useful for identifying a hepatotoxic agent. The present  
CC sequence is given in the exemplification of the present invention.  
XX  
SQ Sequence 1017 BP; 245 A; 245 C; 216 G; 311 T; 0 other;  
  
Query Match 80.3%; Score 1017; DB 22; Length 1017;  
Best Local Similarity 100.0%; Pred. No. 1e-244;  
Matches 1017; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 245 GAACCTCAGACACCCAAAGAGTACCGAAGCATGTGTCCGCTGGGTAGTCTAGAGGG 304  
DB 1017 GAACCTCAGACACCCAAAGAGTACCGAAGCATGTGTCCGCTGGGTAGTCTAGAGGG 958  
  
QY 305 GCGGCATCATCATGACATGCTGTGTCTCTGCGCAAGACAGCATGTTCAGAAATTC 364  
DB 957 GCGGCATCATCATGACATGCTGTGTCTCTGCGCAAGACAGCATGTTCAGAAATTC 898  
  
QY 365 TAAATAGTTTAAATAGTTTAAAGCCGACGACGATTTTACACCCAGTTACTAGAAA 424  
DB 897 TAAATAGTTTAAATAGTTTAAAGCCGACGACGATTTTACACCCAGTTACTAGAAA 838  
  
QY 425 ACGAAGGGAACACATGATGACATGATGATGATGATGATGATGATGATGATGATGAT 484  
DB 837 ACGAAGGGAACACATGATGACATGATGATGATGATGATGATGATGATGATGATGAT 778  
  
QY 485 ATCTACCAAAAAATCTCCGATCATATTCAGAAAGATCTTATAGTACAGTACAT 544  
DB 777 ATCTACCAAAAAATCTCCGATCATATTCAGAAAGATCTTATAGTACAGTACAT 718  
  
QY 545 ATTGCTGCTTAAAGAGGGGCTCTTAAAGAAAGACATTTGCTAAGTTAGCACTGTGAGA 604  
DB 717 ATTGCTGCTTAAAGAGGGGCTCTTAAAGAAAGACATTTGCTAAGTTAGCACTGTGAGA 658  
  
QY 605 TGGCCAGTTTAAATATGAGACTCAAGCCCATCTGGGGAGGAGACAGCGGGGAAGGG 664  
DB 657 TGGCCAGTTTAAATATGAGACTCAAGCCCATCTGGGGAGGAGACAGCGGGGAAGGG 598  
  
QY 665 GCTCAAGAGACACTGATTAAGATGCGCATTTGTCTATCTACTGTTTACAGAAATTAAC 724  
DB 597 GCTCAAGAGACACTGATTAAGATGCGCATTTGTCTATCTACTGTTTACAGAAATTAAC 538  
  
QY 725 CGTTAAAGAGTTTAAAGAGTTTAAAGAGTTTAAAGAGTTTAAAGAGTTTAAAGAGTT 784  
DB 537 CGTTAAAGAGTTTAAAGAGTTTAAAGAGTTTAAAGAGTTTAAAGAGTTTAAAGAGTT 478  
  
QY 785 TGTAAATTAATCTCTACTTATTAAGTCAAAATATCTGTCTCTCTTATGATGACGTCG 844  
DB 477 TGTAAATTAATCTCTACTTATTAAGTCAAAATATCTGTCTCTCTTATGATGACGTCG 418  
  
QY 845 TGTTCACACACTCCACCACACACCCAGACTAGAGAAAGAAATTAATCTTGTAGAGGCA 904

DB 417 TGTTCACACACTCCACCACACACCCAGACTAGAGAAAGAAATTAATCTTGTAGAGGCA 358  
QY 905 ACACAGAGCCAGACTTCTGTTCAAAAGCTGAGAAAGCCGCTGACGCTGTTTATAGAGA 964  
DB 357 ACACAGAGCCAGACTTCTGTTCAAAAGCTGAGAAAGCCGCTGACGCTGTTTATAGAGA 298  
QY 965 ACTCACTATGAATCAAGAGACAGAGCTGTTTACACCCATCTGACGCTACATCAAAAGTT 1024  
DB 297 ACTCACTATGAATCAAGAGACAGAGCTGTTTACACCCATCTGACGCTACATCAAAAGTT 238  
QY 1025 ACGTAAATGACATGGGCTGATTAAGTACAGGTGCTTACATGACGAGCTGCTATTAGGA 1084  
DB 237 ACGTAAATGACATGGGCTGATTAAGTACAGGTGCTTACATGACGAGCTGCTATTAGGA 178  
QY 1085 GGCCTGTGTGTCTACACGAGTCTGGAGCTTACGAGGCTGACCCCTAGACCCAGAA 1144  
DB 177 GGCCTGTGTGTCTACACGAGTCTGGAGCTTACGAGGCTGACCCCTAGACCCAGAA 118  
QY 1145 GCTGAGCTTCTTAAAGACAAAGCTCTCAACAGCTTACGCTTACGCTTCTCAGCAC 1204  
DB 117 GCTGAGCTTCTTAAAGACAAAGCTCTCAACAGCTTACGCTTACGCTTCTCAGCAC 58  
QY 1205 AACGCACTTATGTTTCAAGAGTATTTGGCAATTTCTTAATCTGACCAAGAAATAGGG 1261  
DB 57 AACGCACTTATGTTTCAAGAGTATTTGGCAATTTCTTAATCTGACCAAGAAATAGGG 1

RESULT 3  
AAFI8236/C  
ID AAFI8236 standard; DNA; 2051 BP.  
XX  
XX AAFI8236;  
DT 14-MAR-2001 (first entry)  
XX  
XX Lung cancer associated polynucleotide sequence SEQ ID 255.  
DE  
XX Human; Lung cancer associated protein; neuroprotective; cytosolic;  
KW cardioactive; immunomodulatory; muscular active; vulnerary;  
KW gastrointestinal; nephrotoxic; anti-infective; gynecological;  
KW antibacterial; diagnosis; neural disorder; immune disorder; reproductive;  
KW proliferative disorder; wound healing; infectious disease; ds.  
XX  
OS Homo sapiens.  
PN WO20005180-A2.  
XX  
XX 21-SEP-2000.  
XX  
XX 08-MAR-2000; 2000WO-US05918.  
XX  
XX 12-MAR-1999; 99US-0124270.  
XX  
XX (HUMA-) HUMAN GENOME SCI INC.  
XX (ROSE/) ROSEN C A.  
XX  
XX Ruben SM;  
XX  
XX WP1: 2000-587514/55.  
DR P-PSDB: AAB58360.  
XX  
XX Lung cancer associated gene sequences, referred to as lung cancer  
PT antigens, useful for treatment, prevention, and diagnosis of disorders  
PT such as lung cancer -  
XX  
XX  
XX Claim 1; Page 716-717; 1425pp; English.  
XX  
XX Polynucleotide sequences AAFI7982 - AAFI8424 encode human lung cancer  
CC associated proteins represented in AAB58106 - AAB58548. Lung cancer  
CC associated proteins and polynucleotide sequences, their agonists, and  
CC antagonists may have neuroprotective; cytosolic; cardioactive;  
CC immunomodulatory; muscular active general; vulnerary; gastrointestinal









XX Gene; liver cancer; ds; hepatocellular carcinoma; hepatotropic;  
 KW metastatic liver tumour; cytostatic; expression profile; disease state;  
 KW disease progression; drug toxicity; drug efficacy; drug metabolism.  
 OS Homo sapiens.  
 XX WO200229103-A2.  
 XX 11-APR-2002.  
 XX 02-OCT-2001; 2001WO-US30589.  
 XX 02-OCT-2000; 2000US-237054P.  
 XX (GENE-) GENE LOGIC INC.  
 XX Horne D, Alvares C, Peres-Da-Silva S, Vockley JG;  
 DR WPI; 2002-426119/45.  
 XX  
 PT Diagnosing and detecting the progression of liver cancer.  
 PT hepatocellular carcinoma or metastatic liver tumor in a patient.  
 PT involves detecting the level of expression of two or more genes in a  
 PT liver tissue sample  
 PS Claim 1: SEQ ID NO 1632; 298bp; English.  
 XX  
 CC The invention relates to a novel method for diagnosing and detecting the  
 CC progression of liver cancer, hepatocellular carcinoma or metastatic liver  
 CC tumor in a patient, and differentiating metastatic liver cancer from  
 CC hepatocellular carcinoma in a patient, involving detecting the level of  
 CC expression of two or more genes represented in ABN93503-ABN97455 in a  
 CC tissue sample. The method of the invention has hepatotropic, and  
 CC cytostatic activity. The method is useful for diagnosing and detecting  
 CC the progression of liver cancer, hepatocellular carcinoma and metastatic  
 CC liver carcinoma in a patient. The method is useful for identifying  
 CC expression profiles which serve as useful diagnostic markers as well as  
 CC markers that can be used to monitor disease states, disease progression,  
 CC drug toxicity, drug efficacy and drug metabolism.  
 CC Note: The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from WIFO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.  
 XX  
 SQ Sequence 1232 BP; 329 A; 240 C; 251 G; 411 T; 1 other;

Query Match 33.2%; Score 420.2; DB 24; Length 1232;  
 Best Local Similarity 67.5%; Pred. No. 3.3e-95;  
 Matches 848; Conservative 1; Mismatches 324; Indels 84; Gaps 15;

53 ATTCAAGATTATTATTAAGTCAATGCAAAACATATCTGTAATGCAATTAGCAAAAGAT 112  
 1228 WTCCAGATTTATTATTAAGTCAATGCAAAACATATGCTAATGCAAAAGAT 1169  
 113 CAATGTAATAAACTCCACAAATTCGCAACTGTCATTTGAAAAAGTTGTTAGTGGT 172  
 1168 CAATGTAATAAACTCCACAAATTCGCAACTGTCATTTGAAAAAGTTGTTAGTGGT 1109  
 173 CGAAAGGCCCAACACTGTTCTGTCAGTGTAGTGTAGTACGAAGCGGTTAGAC 232  
 1108 TGAAGGGTCCACGCTGATTTCTGCCAGTGAAGTTGTTACGAACATGCTAGGCAC 1049  
 233 TAGCGC--TTGACAGAACCTCAGACCCAAAG-----TACCG 269  
 1048 TAGCAGAGTTTACAGAACCTCAGACCCAAAGAACATCATATAGCAAAAGGACTATACG 989  
 270 GAAGCATGTGTCGGGTGGGTGAGGTCTAGAGGGGCGGCATCAATCAGATGAGTGT 329  
 988 GAGGCGTGTGTCGGGTGGGTGAGGTAAAGA-GGGTCAGTATTTGGTCAAGTACAGTGTG 930  
 330 GTTACTCTGGCAAGACAGTGTG-----TTTCAGATATCTTAATAATAG 372  
 929 GGTATATCTGGCAAGACAGTGTGTTAAGAGGTTTATGATTTAAGAAATTAATCTTAATAATAT 870

QY 373 TTTAAAACTGTAAAGCCGACAGGTGATTTCTACACCCAGTACTAGTAAACGAAAGG 432  
 869 TTTAAAACTGTAAAGCCGACAGTGTGATTTCTACACCCAGTACTAGTAAACGAAAGG 810  
 QY 433 AAGCACTAGTCACTGAGTAAAGAGGTGAAGAACGACGACCTTCTACTATCTACCA 492  
 809 AAGCACTATATGAGTCTGAAATA-----AAGTAAACATGCAAGAACCTTCTACTAATCGACA 755  
 QY 493 AAAAATCTCCGAATGATATACGAAGATCTTATAGTACAGTGCACATATGCTCCG 552  
 754 AAAAATCTCCGAATGATATACGAAGATCTTATAGTACAGTGCACATATGCTCCG 696  
 QY 553 TTAAGAGGGGGTCTTAAAGAAAGACCTTCTAGTAAAGTAAAGTAAAGTAAAGTAAAG 612  
 695 GTCAAGAGGGGTCTTAAAGAAAGACCTTCTAGTAAAGTAAAGTAAAGTAAAGTAAAG 636  
 QY 613 TTAAGATGAGTCAACGCGCCATCTGGGAGGACAGCAGGAGGGGAGGGGCTCAAGA 672  
 635 TTAAGATGAGTCAACGCGCCATCTGGGAGGACAGCAGGAGGGGAGGGGCTCAAGA 576  
 QY 673 GAGACATGATTAAGATCGGCCATTTGTATCTACTGTTTGAACAAATTAACCGTTAAA 732  
 575 CTTAAGAAACAACTTCTGATATATACCAACTTCTTCTATCTACTGATTTGACAG 516  
 QY 733 AGCTTTACCCGTCGACACTTTTATGAGTTGAA-----TTACTCCATGTACATGATGTTAA 789  
 515 AAATTAACCTTTTAAAGTTTATGAGTTTATGAGTTTATGAGTTTATGAGTTTATGAGTT 457  
 QY 790 AATTAATCTCTACTGATA-TTATGCAAAATTAAGTGTCTGCTTCTGATGAGTGTGTT 848  
 456 ---AATCTCTCTACTGATA-TTATGCAAAATTAAGTGTCTGCTTCTGATGAGTGTGTT 409  
 QY 849 TCACACTCTCCACCCAGCACACCCAGTACTAGAACAAATTAAGTGTGTTAGGAGCAAC 908  
 408 -CACACACCCACCCAGCACACCCAGTACTAGAACAAATTAAGTGTGTTAGGAGCAAC 352  
 QY 909 AGGACAGAGGTTCTGTTCAAGCCTGTCAGAACCCGTCAGTGTGTTAGGAGCAAC 968  
 351 -----CAGTGTCTGTTCAAGCCTGTCAGAACCCGTCAGTGTGTTAGGAGCAAC 298  
 QY 969 ACTATGAATTAAGACAGCAGAGCTGTACACCCATGTCAGTGTGTTAGGAGCAAC 1028  
 297 CAAGACATGATCAAGAGCTGAGTGTGTCAGTGTGTTAGGAGCAAC 238  
 QY 1029 AATGACATGGGCTGATAGTATAGTGTGTCAGTGTGTTAGGAGCAAC 1088  
 237 CGGGAACAGTGTGTCAGTGTGTCAGTGTGTCAGTGTGTCAGTGTGTCAGTGTGTCAGT 178  
 QY 1089 GTGCTGTGTCAACGCTGTG-----GAGCTACGGGAGGAGGTCGACCC-CCGAGGCC 1140  
 177 GTTGTGCTTACGATCTGAGAACATGATAGAGAGCAAGTGTGTCAGTGTGTCAGTGTGTCAG 118  
 QY 1141 AGAAGCTGAGTCTTCTTAAGGACAAAGTCTCTCAACAGCTTATGCTTACGCTGTTCGA 1200  
 117 AGAAGCTGAGTCTTCTTAAGGACAAAGTCTCTCAACAGCTTATGCTTACGCTGTTCGA 58  
 QY 1201 GCACAGC---GCAACTGTGTCAGAGGATTTGGCAATCTTAATGTAGCAAGA 1254  
 57 GCACAGCAGTGTGTCAGAGGATTTGGCAATCTTAATGTAGCAAGA 1

RESULT 7  
 AAH22399/C  
 ID AAH22399 standard; DNA; 348 BP.  
 XX AAH22399;  
 AC 22-AUG-2001 (first entry)  
 DT Human rac1 gene related nucleotide sequence #3.  
 XX Identification; toxic; hepatotoxic; differential gene expression;  
 KW

KW NSAID; non-steroidal antiinflammatory drug; ds.  
XX Homo sapiens.  
OS  
XX WO200138579-A2.  
PN  
XX  
XX 31-MAY-2001.  
PD  
XX 21-NOV-2000; 2000MO-US32049.  
PF  
XX 22-NOV-1999; 99US-0166923.  
PR 18-FEB-2000; 2000US-0183531.  
PR 20-NOV-2000; 2000US-0717321.  
XX  
XX (CURA-) CURAGEN CORP.  
PI Gould-Rothberg BE, Dipippo VA, Ramsen TM, Gerweln RM;  
XX WPI; 2001-355948/37.  
DR  
XX Screening hepatotoxic agent comprises contacting test cell population  
PT expressing RISKMARKER or INJURYMARKER with agent, comparing expression  
PT with reference population and identifying difference in expression  
PT levels -  
XX  
XX Disclosure; Page 8-9; 76pp; English.  
PS  
XX The present invention describes a method of screening a test agent for  
CC hepatotoxicity. The method comprises: (a) providing a test cell  
CC population comprising a cell capable of expressing one or more nucleic  
CC acid sequences selected from the group consisting of RISKMARKER 1-8  
CC and INJURYMARKER 1-10; (b) contacting the test cell population with a  
CC test agent; (c) measuring expression of one or more of the nucleic  
CC acid sequences in the test cell population; (d) comparing the  
CC expression of the nucleic acid sequence in the test cell population to  
CC the expression of the nucleic acid sequence in a reference cell  
CC population comprising at least one cell whose exposure status to a  
CC hepatotoxic agent is known; and (e) identifying a difference in  
CC expression levels of the RISKMARKER or INJURYMARKER sequences. If  
CC present, in the test cell population and reference cell population.  
CC The method is useful for identifying a hepatotoxic agent. The present  
CC sequence is given in the exemplification of the present invention.  
XX  
XX  
SO Sequence 348 BP; 103 A; 76 C; 67 G; 102 T; 0 other;  
Query Match 27.5%; Score 348; DB 22; Length 348;  
Best Local Similarity 100.0%; Pred. No. 2.5e-77;  
Matches 348; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
10 TTTTTCAGTTCGAAGACATTTTTTTTTTTTATGATTCAGATTTATTA 69  
|||||  
348 TTTTTCAGTTCGAAGACATTTTTTTTTTTTATGATTCAGATTTATTA 289  
70 GTCATCATGCAAAACATAGCTGCTAAGTATGCAAAAGATCAATGTAACACTCC 129  
288 GTCATCATGCAAAACATAGCTGCTAAGTATGCAAAAGATCAATGTAACACTCC 229  
130 ACAATTCGCACTGTCATTTGAAAAAGTTTTCAGTTCGCAAAAGCCCAACACTG 189  
228 ACAATTCGCACTGTCATTTGAAAAAGTTTTCAGTTCGCAAAAGCCCAACACTG 169  
190 TGTTCGCGAGTGTAGTGTGTACAGAAAGCGGTATGACATGCGCTGACAGAAC 249  
168 TGTTCGCGAGTGTAGTGTGTACAGAAAGCGGTATGACATGCGCTGACAGAAC 109  
250 TCACAGACCCAAAGTACCGGAAGCATGTGTCCGCTGGGTGAGTCTAGAGGGGGCGG 309  
108 TCACAGACCCAAAGTACCGGAAGCATGTGTCCGCTGGGTGAGTCTAGAGGGGGCGG 49  
310 ATCAATCATGACAGTGTGTGTCTGTGCAAGAAAGATGATTTTCA 357  
|||||  
48 ATCAATCATGACAGTGTGTGTCTGTGCAAGAAAGATGATTTTCA 1

RESULT 8  
ID AAS65569/c  
ID AAS65569 standard; cDNA; 1318 BP.  
XX  
XX AAS65569;  
AC  
XX 13-FEB-2002 (first entry)  
DT  
XX  
XX DNA encoding novel human diagnostic protein #1373.  
DE  
XX  
XX Human; chromosome mapping; gene mapping; gene therapy; forensic;  
KW food supplement; medical imaging; diagnostic; genetic disorder; ss.  
XX  
XX Homo sapiens.  
OS  
XX  
XX WO200175067-A2.  
PN  
XX  
XX 11-OCT-2001.  
PD  
XX  
XX 30-MAR-2001; 2001MO-US08631.  
PF  
XX  
XX 31-MAR-2000; 2000US-0540217.  
PR 23-AUG-2000; 2000US-0649167.  
PR  
XX  
XX (HYSE-) HYSEQ INC.  
PA  
XX  
XX  
XX Dmanac RT, Liu C, Tang YT;  
PI  
XX  
XX WPI; 2001-639362/73.  
DR  
XX P-PSDB; ABG01382.  
DR  
XX  
XX New isolated polynucleotide and encoded polypeptides, useful in  
PT diagnostics, forensics, gene mapping, identification of mutations  
PT responsible for genetic disorders or other traits and to assess  
PT biodiversity -  
XX  
XX  
PS Claim 1; SEQ ID No 1373; 103pp; English.  
XX  
XX The invention relates to isolated polynucleotide (I) and  
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,  
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
CC and gene mapping, and in recombinant production of (II). The  
CC polynucleotides are also used in diagnostics as expressed sequence tags  
CC for identifying expressed genes. (I) is useful in gene therapy techniques  
CC to restore normal activity of (II) or to treat disease states involving  
CC (II). (II) is useful for generating antibodies against it, detecting or  
CC quantitating a polypeptide in tissue, as molecular weight markers and as  
CC a food supplement. (II) and its binding partners are useful in medical  
CC imaging of sites expressing (II). (I) and (II) are useful for treating  
CC disorders involving aberrant protein expression or biological activity.  
CC The polypeptide and polynucleotide sequences have applications in  
CC diagnostics, forensics, gene mapping, identification of mutations  
CC responsible for genetic disorders or other traits to assess biodiversity  
CC and to produce other types of data and products dependent on DNA and  
CC amino acid sequences. AAS64197-AAS94564 represent novel human  
CC diagnostic coding sequences of the invention.  
CC Note: The sequence data for this patent did not appear in the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
XX  
SO Sequence 1318 BP; 346 A; 268 C; 279 G; 424 T; 1 other;  
Query Match 23.9%; Score 303; DB 23; Length 1318;  
Best Local Similarity 62.3%; Pred. No. 7.6e-66;  
Matches 805; Conservative 1; Mismatches 386; Indels 101; Gaps 17;  
53 ATCAAGGATTTTAAATGATCATGCAAAACATAGCTGATTCATGCAAAAGAT 112  
1314 WTCCAAAGATTTTAAATGATCATGCAAAACATAGCTGATTCATGCAAAAGAT 1255  
|||||  
113 CAATGTAAACACATCTGCAACTGTCATTTGAAAAAGTTTGTCTAGTGGT 172  
|||||

```

Db      1254 CAATGTAAAAACCTCCACAAATTCCTGCACTCTCAATTTAAAAAAATCTGTGTAGTGC 1195
Oy      173  CGAAGGCCCAACACTGTGTCTTCTGCACTAGTAGTGGTTGACAGAGGGGTTAGAC 232
Db      1194 TGAAGGGTCCACGCTGTATCTCCGCACTGAGTTAGTTGACGAACATGTCAGCAC 1135
Oy      233  TAGCGC--TTGACAGAACCTCCACAGACCCAAAG-----TACC 269
Db      1134 TAGCAGACTTTACAGAACCTCCACAGACCCAAAGAACATCAATAGCAAAAGCACTACAG 1075
Oy      270  GAAGCATGTCTCCGGTGGTGGAGTCTAGAGGGGGCGCATCAATCAGTACAGTGT 329
Db      1074 GAGCGGTGTCTCCGGTGGGAGTAAAGA-GGGTCAGTATGTGTCAAGTCAAGTGTCTC 1016
Oy      330  GGTACTCTGGCAAGACAGTGTATGTTC-----GAATATCTAAATATG 372
Db      1015 GGTATCTGGCAAGACAGTGTATGTTCAGAGTGTATGTATAGAAATATCTTAAATAT 956
Oy      373  TTTAAAACTGTAAAGCCGACAGCATGTTCTACACCGTACTAGAAAACGAGAGG 432
Db      955  TTTAAAACTGTAAAGCTGCAACACATGATTTTACACTGTACTAGAAAACGAGAG 896
Oy      433  AAGCACTGTAGCTGAGTAAAGAGGTGAAAACGAGACGACTTCTACTATACCA 492
Db      895  AAGCACTGTAGCTGTGAATAAA-----GTAAACATGGAAGCACTTTACTAATCGACA 841
Oy      493  AAAAAATCTCGAATGATTCATGAGAAAGATCTTATAGACAGTCAAGCATATTTGCTG 552
Db      840  AAAAACTTTATATGATTCATGAGAAAGATTTTATATACAG-GAGGCATATGTGCA 782
Oy      553  TTAAGAGGGGGTCTTAAAGAA-----AGCACTGTCTAAGTTACCAACTGTGAGATGG 607
Db      781  GTCAAGAGGGGTTCTTTTAAAGAAAGCACTTTCTAAGTTAAAGCACTAACAGAACAA 722
Oy      608  CCAAGTTAAATATGACTCAA---CGCCCATCTG66GAGGACACGAGGGGAAAGGGG 664
Db      721  CCAGTTTAAAGCAATTAATTAATGCCCCCAATTTGGGAGGAGCATGGGAGGTAAAGGA 662
Oy      665  GCTCAAGAGA-----GACACTGATTAAGTCGGCCATTTGTCTAT 702
Db      661  AAGGAAAAGGCTTAAGAAAACATTTCTTGATTAATCCMAACCTTTCTTCAATTATCTA 602
Oy      703  CTACTGTTTGAACAGAAATTAACCGTTAAAGCTTTTACCGGTGACACTTTTATTCAGTGG 762
Db      601  CTGCAITGGACCAAGAAATTAACCTTTTAAAGATTTTACCCCGTGACACACTTTCATCC 542
Oy      763  AATTAATCAGATGTACAAATGTAGTGAATTAATCTCTACTCATTTAGTCAAAATACTG 822
Db      541  CTTGTACCAATGTAGTGTGAATTCCTCCACTCCGTATTTTGTTCACA---AAATTAAGTG 485
Oy      823  TCTGTCTCTTGTGATGAGTGTGTTCACACACTCCACCCAGACACCCAGCACTAGGA 882
Db      484  TCTTTGTCTTTGA-----TCCACACACACACCCCGGACACCCGACACACTTAA 430
Oy      883  ACAGAAATCTTCTGTAGAGGACACACAGAGCCAGAGTCTGTGTAAAGCCCTGCAAGAGC 942
Db      429  CCAGAAATTCCTTCAATTAAGGAATAGCCAGTTCGTTCAAATGTCCCGGAAAAGCTGG 370
Oy      943  CGGTACAGTGTATTTTGAAGAACTCACTATGAATCAAAAGACAGAGCTGTACACCA 1002
Db      369  TCAGAAAACCTGCAATGAATACCAAAACATGATCCAAAGAGCTGTGCTGCTGCA 310
Oy      1003  TCGTACAGTACAGTCAAAAGTTAGTATAGCATGGGCTG-ATAAGTTACAGTGGCGTT 1061
Db      309  --CTTCATTAACATGTAATGTATGTGGGAACAGTCTGCGTAACCTACAGTGTGAGTT 252
Oy      1062  ACATGGCAGCGTGTCAATTAAGAGGCTGTGCTGTCTACACAGGCTGTGGGA-----GC 1113
Db      251  CAATGGCAACGCTTCATTCGGGAGGCTGTCTGCTTTACGCACTGAGAACTACATAGGA 192
Oy      1114  TACGGGAGGCTGTGACC-CTTGAGCCAGAAAGCTGCACTTCTTTAA--GGACAAAGTC 1170
Db      191  GAGCAAAAGTGTGTGCACTCTTAACGTGAGAGGCTACGCTCTTCTCAATGACGAGAGTGC 132

```

```

Oy      1171 TCTCAACAGCTTAGTGC--TACGTGTCTCAGCACAAAGCAACTTAGTTCAGAGTAT 1228
Db      131  TTTCGAAAGTTACAGTGTCTCGGTCTCTCGGCACAAACAAATGCACTAGTATTCAAGAGTAT 72
Oy      1229  TTTGGCAATTCCTTAATCTGAGCAAGAAATAGGG 1261
Db      71  TTTGGCAACTCTTAATCTGAACAGAAATAGGGG 39

RESULT 9
AAH232398/C
AAH232398 standard; DNA; 1022 BP.
AC      AAH232398;
XX
XX
XX      22-NOV-2001 (first entry)
DT
XX
XX      Human rac1 gene related nucleotide sequence #2.
DE
XX
XX      Identification: toxic; hepatotoxic; differential gene expression;
KW      NSAID; non-steroidal antiinflammatory drug; ds.
XX
XX      Homo sapiens.
OS
XX      WO200138579-A2.
PN
XX      31-MAY-2001.
PD
XX
XX      21-NOV-2000; 2000WO-0S32049.
PF
XX
XX      22-NOV-1999; 99US-0166923.
PR      18-FEB-2000; 2000US-0183531.
PR      20-NOV-2000; 2000US-0717321.
XX
XX      (CURA-) CURAGEN CORP.
PA
XX
XX      Gould-Rothberg BE, DiIppio VA, Ramseah TM, Gerwehn RM;
PI      WPI: 2001-355948/37.
XX
XX      Screening hepatotoxic agent comprises contacting test cell population
PT      expressing RISKMARKER or INJURYMARKER with agent, comparing expression
PT      with reference population and identifying difference in expression
PT      levels -
XX
XX      Disclosure: Page 7-8; 76pp; English.
PS
XX
XX      The present invention describes a method of screening a test agent for
CC      hepatotoxicity. The method comprises: (a) providing a test cell
CC      population comprising a cell capable of expressing one or more nucleic
CC      acid sequences selected from the group consisting of RISKMARKER 1-8
CC      and INJURYMARKER 1-10; (b) contacting the test cell population with a
CC      test agent; (c) measuring expression of one or more of the nucleic
CC      acid sequences in the test cell population; (d) comparing the
CC      expression of the nucleic acid sequence in the test cell population to
CC      the expression of the nucleic acid sequence in an reference cell
CC      population comprising at least one cell whose exposure status to a
CC      hepatotoxic agent is known; and (e) identifying a difference in
CC      expression levels of the RISKMARKER or INJURYMARKER sequences, if
CC      present, in the test cell population and reference cell population.
CC      The method is useful for identifying a hepatotoxic agent. The present
CC      sequence is given in the exemplification of the present invention.
XX
XX      Sequence 1022 BP; 273 A; 208 C; 205 G; 336 T; 0 other;
SQ

Query Match      23.0%; Score 290.6; DB 22; Length 1022;
Best Local Similarity 64.5%; Pred. No. 8,8e-63;
Matches 664; Conservative 0; Mismatches 304; Indels 61; Gaps 13;
Oy      265  TACGGAAGCATGTGTCCGCGTGGTGAAGTCTAGAGGGGGGCGCATCAATCACATGACA 324
Db      1000  TACAGAGAGCGGTGTCCGCGTGGCGGAGGTAAAGA-GGTCACATTTGTGTCAAGTGA 942

```

QY	325	GTGTTGGTACTCTGGCAAGACGAGATGTTTCACAA-----TATCTAA	367
Db	941	GTCTCGGTATCTCTGGCAAGACGATGTTTAAAGAGTCTACTAGTTTAAAGATTATCTAA	882
QY	368	AATAGTTTAAAAACTGTAAAGCCGACAGACGTGATTTCTACCCAGTACTAGAAAAAG	427
Db	881	AATATTTTAAAAACTATTAAGCTGCACACATGATTTTAAACCGTAGTACTAGAAAAGT	822
QY	428	AAGGAAGCACTACTCAGCTGATGTTAAAGGAAGGTGMAAACGGAACGACTTCTACTATC	487
Db	821	AAGGAAGCACTTATTAAGCTCTGAAATA---AAGTAACATGGAAGCACTTTTACTATAT	767
QY	488	TACCAAAAAAAATCTCCGAATGCTATTATCAGAAAGATCTTATAGTACAGCTGACATAT	547
Db	766	CGACAAAAAAACCTTCTATATCATTTATCAGAAAGTTTATATATATCAAG-GAGGCATAT	708
QY	548	GCTCGTTTAAAGAGGGGGTCTTAAAGAAAAAGCACTTGTCAAGTTAGCAACTGTGAGATG	607
Db	707	GCTCAGTCAGAAAGGGGTCTTATTAAGAAAAACCACTTACTTAAGTTAGGCACTTAACGAACAA	648
QY	608	CCACTTTTAAATATGACTCAACGCCCTCTGGGGAAGGGAACAGCAAGGGGGAAGGGGGCT	667
Db	647	CCAGTTTAAACATTAATTAATTAATGCCCCAATTTGGGGAAGGCAAGGAGGTGAAGGAAGAG	588
QY	668	CAGAAGACACTATATAGATCGGCCATTTGTCACTCTCTGTTGACAGAAATTAACCGT	727
Db	587	AAAAGCTTAAACAAAACATTTCTGTATATATACCAACCTTTCTTCACTATCTACGCAAT	528
QY	728	TAAAAAGCTTTACCCGTGACACTTTTATTCAGTTGAAT---ACTCCATGTACAATGTAG	784
Db	527	GACGAAATTAACCTTTAGAGTTTTCACCGGTACACTTTCATTTCCCTTGACAAATGTAG	468
QY	785	TGTAAATTAATCTGACTTCAATA-TTGTCAAAATACGTGCTGCTCCTTGTATGAGCTC	843
Db	467	TGTA---AATCTCACCTCTGATTTTGTCTCAAAATACGTGTTGTCTCTTGATCA----	416
QY	844	GTGTTTTCACACTCCACCCAGACACCCAGACAGTATGAGAACAGATACTTGTTAGAGC	903
Db	415	-----CACACACCCACCCGGACACCCACAGCTA--AACAGATTTCTTATTAGAGGA	364
QY	904	AACACAGAGCCAGAGATTTCTTTCAAGCCCTGCAGAACCCGGTCACTGTTATTTAGAG	963
Db	363	AATA-----GCAATCTGTTCAAAAATCTCCGCAAAACCTGTCAGAAAACCTCGCTATGA	310
QY	964	AATCACTATGAATTCAAAGAGACGAGCTGTTAACCCATGTCAGTCATAGTACAAGT	1023
Db	309	AATCAACAAGACTGATCCAAAGAGCTGAGTCTGCTACCTCCTCATTAAGTACAAGT	250
QY	1024	TACGTAATGACATAGGGCTGATTAAGTTACAGTGGCTGTACATGGCAGCGTGTCAATTAAG	1083
Db	249	TATGTGCGGAACAGTGTGCTTAOTCACTGCTGTGAGTTCATAGGGAACGTTCAATTCGG	190
QY	1084	AGGCTGTCTGTGTACACGCTCTGG-----GAGCTACGGGAAGGCTCTGCACG-CCTG	1133
Db	189	AGGCTGTCTGTCTTACGCACTGTAGAACTACATATAGGAGACAAAGTCTCACCTCCTTA	130
QY	1136	AGCCACAGAAAGTGTGAGTCTCTTAAAGCAAGTCTCCAAAGCTTGTAGTGTGCTTACGAT	1195
Db	129	ACTGCAGAGCTACCGTCTTCTCAAAAGCAGAAAGCTTTGCAGAAAGTTAGTGTCTCGGTGT	70
QY	1196	TCTCAGCACAAC---GCAACTTAGTTCACAAGATATTTTGGCAATTTCTTAATCTGAGCAA	1253
Db	69	TCTCGGCACAACAATGACGTAGTGTTCAGAAAGATTTTGGCACTCTTAATCTGAAACA	10
QY	1253	GAATAGGGG 1261	
Db	9	GAATGGGG 1	

XX	AAA89693:	
XX	08-JAN-2001	(first entry)
XX		
DE	Mouse Exo103 nucleotide sequence.	
KM	Exocytosis pathway protein; Exo protein; antiparkinsonian; antidiabetic;	
KM	anti-allergic; antispasmodic; nootropic; neuroprotective; anticonvulsant;	
KM	vulnerary; asthma; inflammation; allergy; Chediak-Higashi syndrome; CHS;	
KM	Alzheimer's disease; Parkinson's disease; Huntington's disease; diabetes;	
XX	digestion disorder; wound healing disorder; gene therapy; ss.	
OS	Mus sp.	
XX		
PN	MO200043419-A2.	
XX		
PD	27-JUL-2000.	
XX		
PF	20-JAN-2000; 2000MO-US01431.	
XX		
PR	20-JAN-1999; 99US-0116534.	
PR	26-JAN-1999; 99US-0117274.	
PR	26-JAN-1999; 99US-0117308.	
PR	26-JAN-1999; 99US-0117309.	
PR	26-JAN-1999; 99US-0117312.	
PR	01-FEB-1999; 99US-0118177.	
PR	01-FEB-1999; 99US-0118178.	
PR	01-FEB-1999; 99US-0118179.	
PR	09-FEB-1999; 99US-0119286.	
PR	11-FEB-1999; 99US-0119998.	
PR	11-FEB-1999; 99US-0119759.	
PA	(RIGE-) RIGEL PHARM INC.	
XX		
PI	Luo Y;	
XX		
DR	WPI: 2000-482908/42.	
XX		
PT	New nucleic acids encoding Exo proteins which are useful in the	
PT	diagnosis, treatment or prevention of exocytosis-mediated disorders	
PT	such as asthma, inflammation and allergies -	
XX		
PS	Disclosure; Page 271; 305pp; English.	
XX		
CC	The present sequence encodes a polypeptide which is associated with	
CC	the exocytosis pathway. cDNA molecules encoding proteins involved in	
CC	exocytosis have been isolated by yeast one-hybrid and two-hybrid	
CC	screening. Novel proteins, termed Exo proteins, have been identified that	
CC	interact with known exocytosis-associated proteins such as GS27, alpha	
CC	snap, unc18-1, vamps3, snap23, and the rab family of proteins.	
CC	Exo proteins and their agonists and antagonists are useful in the	
CC	diagnosis, treatment or prevention of exocytosis-mediated disorders	
CC	such as asthma, inflammation, allergies, Chediak-Higashi Syndrome	
CC	(CHS), Alzheimer's disease, Parkinson's disease, Huntington's disease,	
CC	diabetes, digestion disorders and wound healing disorders.	
CC	The nucleic acids, antagonists or agonists of Exo proteins are useful	
CC	in gene therapy. The nucleic acids are also useful for generating	
CC	transgenic or knock-out animals which can be used in the	
CC	development and screening of therapeutically useful reagents.	
XX		
SO	Sequence 447 BP; 102 A; 95 C; 93 G; 157 T; 0 other;	
	Query Match 18.2%; Score 230.6; DB 21; Length 447;	
	Best Local Similarity 78.4%; Pred. No. 7e-48;	
	Matches 349; Conservative 0; Mismatches 69; Indels 27; Gaps 5	
Oy	77 ATGCAAAACATCTGCTAATGATTCAGTAAAGATCAATTAACACATCCACAATTC 136	
Db	447 ATGCAAAACCTACTGCTAATGATTCAGTAAAGATCAATTC AAAACACTCCACAATTC 389	
Oy	137 TGGCACTGTCATTCATGAAAAAGCTTTGTC-TAGTGGTCGAAAGCCCAACACTGTGTC 195	

DB 388 TGCACCTGCAATTCAGAAAAAGTTGTTCTTACGTGTTGAAAGCCCACTGCAATTC 329  
QY 196 TGCACGATGATGATGTTGTACAGAAAGCGCTTACGACCTAGCGCTTACAGAACTTCACAG 255  
DB 328 TGCCAGTGTAGTGTAGTTAGACAGACGCTTACGACCTAGCGCTTACAGAACTTCACAG 269  
QY 256 ACCCAAG-----TACCGAAGCATGTGTCCGCTGGGTAGG 294  
DB 268 ACCCAAGCAATCTACGACAGACCCATGACAGACCGCTGTCCACGTGGCGAGG 209  
QY 295 TCTAGAGGGGGCGCATCATCATGACAGCTGTGTGACCTGTGCAACAGCATATGT 354  
DB 208 TCTA---GGAGATTAATGTCACATGACAGCTGTGTGTAGAAAGTTTCATATG 152  
QY 355 TCAGATATCTAAATAGTTTAAACCTGTAAAGCCGACGACGATTTTACACCCAG 414  
DB 151 TAAGAAATCTAAATATTTTAAACCGTTAAAGCTGCACACATGATTTGCACACCTAG 92  
QY 415 TTACTAGAAAGCAAGGACGACCTAGT-CAGCTGATGTAAGGAAAGTGAAGAAACAGAAC 473  
DB 91 TTGCTAGCAAACTAAGGACGACCTAAGCTGTGAGTAAAGTAAAGCGAAACAGACG 32  
QY 474 GCACTTCTACTATCTACCAAAAAA 498  
DB 31 CACTTCTACTAGCTACCAAAAAA 7

## RESULT 11

AAA89694/C  
ID AAA89694 standard; cDNA: 422 BP.

AC AAA89694;

DT 08-JAN-2001 (first entry)

XX Mouse Rab2 nucleotide sequence #1.

XX Exocytosis pathway protein; Exo protein; antiparkinsonian; antidiabetic;  
XX antiasthmatic; neuroprotective; anticonvulsant;  
XX vulnerability; asthma; inflammation; allergy; Chediak-Higashi syndrome; CHS;  
XX Alzheimer's disease; Parkinson's disease; Huntington's disease; diabetes;  
XX digestion disorder; wound healing disorder; gene therapy; ss.

OS Mus sp.

PN WO200043419-A2.

PD 27-JUL-2000.

XX 20-JAN-2000; 2000WO-US01431.

PF 20-JAN-1999; 99US-0116534.

PR 26-JAN-1999; 99US-0117274.

PR 26-JAN-1999; 99US-0117308.

PR 26-JAN-1999; 99US-0117309.

PR 01-FEB-1999; 99US-0118177.

PR 01-FEB-1999; 99US-0118178.

PR 01-FEB-1999; 99US-0118179.

PR 09-FEB-1999; 99US-0119286.

PR 11-FEB-1999; 99US-0119998.

PR 11-FEB-1999; 99US-0119759.

XX (RIGEL-) RIGEL PHARM INC.  
XX Luo Y.  
XX WPI: 2000-482908/42.  
XX New nucleic acids encoding Exo proteins which are useful in the  
XX diagnosis, treatment or prevention of exocytosis-mediated disorders  
XX such as asthma, inflammation and allergies -

PS Disclosure: Page 271-272; 305pp; English.  
XX  
CC The present sequence encodes a polypeptide which is associated with  
CC the exocytosis pathway. cDNA molecules encoding proteins involved in  
CC exocytosis have been isolated by yeast one-hybrid and two-hybrid  
CC screening. Novel proteins, termed Exo proteins, have been identified that  
CC interact with known exocytosis-associated proteins such as GS27, alpha  
CC snap, unc18-1, vamps, snap-23, and the rab family of proteins.  
CC Exo proteins and their agonists and antagonists are useful in the  
CC diagnosis, treatment or prevention of exocytosis-mediated disorders  
CC such as asthma, inflammation, allergies, Chediak-Higashi Syndrome  
CC (CHS), Alzheimer's disease, Parkinson's disease, Huntington's disease,  
CC diabetes, digestion disorders and wound healing disorders.  
CC The nucleic acids, antagonists or agonists of Exo proteins are useful  
CC in gene therapy. The nucleic acids are also useful for generating  
CC transgenic or knock-out animals which can be used in the  
CC development and screening of therapeutically useful reagents.  
XX  
SQ Sequence 422 BP; 95 A; 92 C; 83 G; 148 T; 4 other;

Query Match 17.0%; Score 215.8; DB 21; Length 422;

Best Local Similarity 77.0%; Pred No. 3.5e-44;

Matches 322; Conservative 0; Mismatches 70; Indels 26; Gaps 4;

QY 104 GCMAAAGATCAATGTAAAAAAGCT-CCACATCTGCAACTGTCAATTTGAAAAAGTTTG 162

DB 421 GCMAAAGATCAATGTAAAAAAGCTCCCAATTTGCAACGTCAATTTGAAAAAGTTTG 362

QY 163 TTCTAGTGTGCAAAAGGCCCAACTGTGTTTCCAGTGAGTTAGTTGTACGAACG 222

DB 361 TTCTAGTGTGCAAAAGGCCCAACTGTGTTTCCAGTGAGTTAGTTGTACGAACG 302

QY 223 GCGTTAGCCTAGCGCTGACAGAACCTCAGACGACCAAGS----- 264

DB 301 GCGTTAGCCTAGCAGTTGTACAGAACCTCAGACGACCAAGCAATCTTTAGCGAGAC 242

QY 265 ---TACCGAAGCATGTGTCCGCTGGGTGAGTTAGAGGGGGCGCATCATCATG 321

DB 241 CATGACAGGAAGCGCTGTCCACGTCGGCGAGGTTAG---GCCAGCATTTAGTCACATG 185

QY 322 ACAGTGTGTGCTCTGTGCAAGACAGTATGTTTGAAGATCTTAAATAGTTTAAAC 381

DB 184 ACAGTGTGTGCTCTGTGCAAGACAGTATGTTTGAAGATCTTAAATAGTTTAAAC 125

QY 382 TGTAAAGCCGACGATGATTTTACACCCAGTTACTAGAAAGCAAGGAAAGCACTAG 441

DB 124 CGTAAAGCTGCAACACATGATTTGCAACACCTAGTTGCTAGGAACCTAAAGAACCTAAC 65

QY 442 T-CAGCTGATTAAGGAAGGTGAAAGCAAGAACGCACTTCTACTATCTACCAAAAAA 498

DB 64 TAGCTCTGATTAAGTAAAGGCGAAAGCAAGACGCACTTTTACTAGCTACCAAAAAA 7

## RESULT 12

ABL38065  
ID ABL38065 standard; cDNA: 605 BP.

AC ABL38065;

DT 08-APR-2002 (first entry)

XX Human colon tumour antigen polynucleotide seq ID NO:1654.

XX Human colon tumour antigen; cytostatic; vaccine;

XX Human colon tumour metastatic antigen; diagnosis; gene; ss.

XX Homo sapiens.

XX WO200196388-A2.

XX 20-DEC-2001.

XX 08-JUN-2001; 2001WO-US18557.

XX 09-JUN-2000; 2000US-210899P.  
 PR 20-FEB-2001; 2001US-270216P.  
 XX  
 PA (CORI-) CORIXA CORP.  
 XX  
 PI Jiang Y, Harlocker SL, Secret H;  
 XX  
 DR WPI: 2002-114514/15.  
 XX  
 PT Novel isolated colon tumor polynucleotide differentially expressed in  
 PT colon tumor or colon metastatic tumor and polypeptides encoded by them,  
 PT useful for inhibiting development of cancer in patient -  
 XX  
 PS claim 1; SEQ ID 1654; 105pp; English.  
 CC ABL36412 to ABL38645 represent human colon tumour antigen cDNA clones (1)  
 CC which were isolated from human colon tumour and colon metastatic tumour  
 CC CDNA libraries. (1) have cytoskeletal activity and can be used in vaccine  
 CC production. (1) can be used for stimulating and/or expanding T cells  
 CC specific for a tumour protein on contact with the T cells. They are also  
 CC useful for inhibiting the development of cancer in a patient. (1) can be  
 CC used as probes or primers for nucleic acid hybridisation, for preparing  
 CC mutant species primers, or primers for use in genetic constructions. (1)  
 CC can be used in the diagnosis of a colon tumour.  
 XX

Sequence 605 BP; 219 A; 109 C; 117 G; 156 T; 4 other:

Query Match 15.3%; Score 193.2; DB 24; Length 605;  
 Best Local Similarity 70.5%; Pred. No. 1.8e-38;

Matches 371; Conservative 0; Mismatches 115; Indels 40; Gaps 7;

265 TACCGAGCATGTGTCGCCGTGGGTGAGTCTAGAGGGGGCGCATCATCAGACA 324  
 74 TACAGAGGGGGTGTGCCGGTGGGAGGTAAAGA-GGGTCAGTATGTGTCAGAGGCCA 132  
 325 GTGTGGTACTCTGGCAGACAGTGAATGTTTCAGAA-----TATCTTA 367  
 133 GTTCGTATCTGGCAGACAGTGAATGTTTCAGAA-----TATCTTA 192  
 368 AATAGTTTAAAGCTGTAAGCGCAGCAGTATTTCTACCCAGTTACTGAAACG 427  
 193 AATATTTTAAAGCTGTAAGCGCAGCAGTATTTCTACCCAGTTACTGAAACG 252  
 428 AAGCGAAGCAGTGAAGCGCAGCAGTATTTCTACCCAGTTACTGAAACG 487  
 253 AAGCGAAGCAGTGAAGCGCAGCAGTATTTCTACCCAGTTACTGAAACG 307  
 488 TACCAAAATATCTCGAATGATATCAGAAAGATTTTATAGTACAGTCAATAT 547  
 308 CGACAAAATATCTCGAATGATATCAGAAAGATTTTATAGTACAGTCAATAT 366  
 548 GTCGTGTTAGAGAGGGGGTCTTAAGAAAGCAGTCTAGTAAAGTGAAGATG 607  
 367 GTCATGATCAGAAAGGGGTTCTATTAAGAAAGCAGTCTAGTAAAGTGAAGATG 426  
 608 CCAATTTAATATGAGTCAAGCCCATCTGGGAGGAGCAGCAGG-----GGG 657  
 427 CCAATTTAATATGAGTCAAGCCCATCTGGGAGGAGCAGCAGG-----GGG 486  
 658 AAGGGGGGTCAGAGAGCAGTGAATGATGCGC-----CATTTGTCATCTAGT 711  
 487 AAAAGCTTAAGAAATCTTCGATATATCAACCTTTCTTCATCATCTACTGATTT 546  
 712 GACAGAAATTAACCGTTAAAGCTTACCCTGACACTTTATTC 757  
 547 GACAGAAATTAACCGTTTACAGNTTTTACCCNTGACACTTTCATTC 592

RESULT 13  
 AAH22400/c  
 ID AAH22400 standard; DNA; 352 BP.

AAH22400;  
 22-AUG-2001 (first entry)  
 Human rac1 gene related nucleotide sequence #4.  
 Identification; toxic; hepatotoxic; differential gene expression;  
 NSAID; non-steroidal antiinflammatory drug; ds.  
 Homo sapiens.  
 W0200136579-A2.  
 31-MAY-2001.  
 21-NOV-2000; 2000MO-US32049.  
 22-NOV-1999; 99US-0166923.  
 18-FEB-2000; 2000US-0183531.  
 20-NOV-2000; 2000US-0717321.  
 (CURA-) CURAGEN CORP.  
 Gould-Rothberg BE, Dipippo VA, Ramsen TM, Gerwein RW;  
 WPI: 2001-355948/37.

Screening hepatotoxic agent comprises contacting test cell population  
 expressing RISKMARKER or INJURYMARKER with agent, comparing expression  
 PT with reference population and identifying difference in expression  
 PT levels -  
 XX  
 PS Disclosure; Page 8-9; 76pp; English.

The present invention describes a method of screening a test agent for  
 hepatotoxicity. The method comprises: (a) providing a test cell  
 CC population comprising a cell capable of expressing one or more nucleic  
 CC acid sequences selected from the group consisting of RISKMARKER 1-8  
 CC and INJURYMARKER 1-10; (b) contacting the test cell population with a  
 CC test agent; (c) measuring expression of one or more of the nucleic  
 CC acid sequences in the test cell population; (d) comparing the  
 CC expression of the nucleic acid sequence in the test cell population to  
 CC the expression of the nucleic acid sequence in a reference cell  
 CC population comprising at least one cell whose exposure status to a  
 CC hepatotoxic agent is known; and (e) identifying a difference in  
 CC expression levels of the RISKMARKER or INJURYMARKER sequences, if  
 CC present, in the test cell population and reference cell population.  
 CC The method is useful for identifying a hepatotoxic agent. The present  
 CC sequence is given in the exemplification of the present invention.

Sequence 352 BP; 86 A; 77 C; 72 G; 117 T; 0 other:

Query Match 14.8%; Score 187.8; DB 22; Length 352;  
 Best Local Similarity 77.6%; Pred. No. 3.4e-37;

Matches 274; Conservative 0; Mismatches 57; Indels 22; Gaps 3;

10 TTTTTCAGAGTCCAAAGACATTTTTCATTTTTCATTTTCATTTTCATTTTCATTT 69  
 352 TTTTTCAGAGTCCAAAGACATTTTTCATTTTTCATTTTTCATTTTCATTTTCATTT 293  
 70 GTATATCATGCAAAAGCTACTGATGATGATGATGATGATGATGATGATGATGATGAT 129  
 292 GTATATCATGCAAAAGCTACTGATGATGATGATGATGATGATGATGATGATGATGAT 233  
 130 ACAATTCGCACTGTCATTTGAAAGATTTTTCATTTTCATTTTCATTTTCATTTTCATTT 189  
 232 ACAATTCGCACTGTCATTTGAAAGATTTTTCATTTTCATTTTCATTTTCATTTTCATTT 173  
 190 TGTTCGCACTGTCATTTGAAAGATTTTTCATTTTCATTTTCATTTTCATTTTCATTT 247  
 172 TATTCGCACTGTCATTTGAAAGATTTTTCATTTTCATTTTCATTTTCATTTTCATTT 113  
 248 CTCACAGACGCCAAAGGTAC-----CGGAAGCATGTGTGCCGGTGG 288



Db 133 GTGTCGTAATCTGGCAAGACAGTGATGTTAAGAAGTTCATAGTTAAGAATTATCTAA 192

CC The invention relates to a purified nucleic acid molecule associated with lactation or muscle and fat deposition (designated IWN) derived

The invention relates to a purified nucleic acid molecule associated with lactation or muscle and fat deposition (designated LMFD), derived from cattle, and the LMFD nucleic acid can specifically hybridise to a

CC second nucleic acid molecule comprising any of 15112 nucleotide  
CC sequences, appearing as ABX34836-ABX49947, or complements of them.  
CC Also included are: (1) a transformed cell having a nucleic acid  
CC comprising an LMF nucleic acid linked to a promoter and a 3' non-  
CC translated sequence that functions in the cell to cause termination of  
CC transcription and addition of polyadenylated ribonucleotides to a 3' end  
CC of the mRNA molecule; and (2) determining a level or pattern of a  
CC molecule in a bovine cell or tissue comprising: (a) incubating a marker  
CC nucleic acid (comprising any of the 15112 nucleic acid sequences or its  
CC complement or fragment) with a complementary nucleic acid molecule  
CC obtained from the bovine cell or tissue, where hybridisation between the  
CC marker nucleic acid and the complementary nucleic acid permits the  
CC detection of the molecule; and (b) detecting the level or pattern of the  
CC complementary nucleic acid, where the detection of the complementary  
CC nucleic acid is predictive of the level or pattern of the molecule.  
CC The LMF nucleic acid is used for determining a level or pattern  
CC of a molecule in a bovine cell or tissue. It is useful for genome  
CC mapping, gene identification and analysis, cattle breeding, preparation  
CC of constructs for use in cattle gene expression, or for genetically  
CC improving cattle. The present sequence is one of the 15112 bovine  
CC LMF EST (expressed sequence tag) nucleic acids.  
CC Note: The present sequence was not shown in the specification but  
CC was obtained in electronic format from the USPTO web site:  
CC seqdata.uspto.gov/sequence.html?docid=20020137139.

XX  
SQ Sequence 432 BP; 123 A; 76 C; 78 G; 155 T; 0 other;

Query Match 13.6%; Score 172.2; DB 25; Length 432;

Best Local Similarity 72.2%; Pred. No. 2.9e-33;

Matches 294; Conservative 0; Mismatches 103; Indels 10; Gaps 5;

QY 379 AACTGTAAGCCGACAGCGATTTCTA-CACCCAGTTACTAGAAACGAGGAAAGCA 437  
Db 432 AACCATTAACCTCGACGCTGATTTTATACCTAGTACTAGAAAACTAAGGAAAGCA 373  
QY 438 CTAGTCAGC--TGAGTAAAGGAAGGTGAAAACAGAGCACTTCTACTATCAAAA 495  
Db 372 CTTATTAGCTTTGATTAACCAACATGAAAACAGAGTGACCTTTTACTACCAAAA 313  
QY 496 AAATCTCCGAATCATTTATGAGAAAGATCTTATAGTACAGGTCAGACATATTCGCTGTTA 555  
Db 312 AAATTTCTTAATGCATTTTCAGAAAGATTTTATTAATACAG-GAGGCATATTCCTCATTA 254  
QY 556 AGAAGGGGGCTCAAGAAAGCACTGCTAAGTTAGCACTGAGAGGATGCCAGTTA 615  
Db 253 AGAATGAGTTCTATTAAGAAAAGCACTTACTAAGTTAGCAACTATAGAGATGACCAAGTCA 194  
QY 616 AATATGAGACTCAAGGCC---CCATCTGGGGAGGAGACAGGAGGGAAGGGGCTCAAG 671  
Db 193 GAGATGATTAATAGCCCAATTTCAGAAAGGGTGGGCCGTTTAAGGAAAAGCTTAAGAAA 134  
QY 672 AGAGACACTGATAGATCGGCCATTTGTCATCTACTG--TTTGACAGAAAATTAAACGTTA 729  
Db 133 ACACCTACTGATATATACCGACCTTCTCATCTACCGCATTTGACAGAAATTAACCTTTT 74  
QY 730 AAAAGCTTACCGGTGACCTTTATTCAGTCAATTAATACCATGTA 776  
Db 73 AAAACTTTTACCCGTGATGCTTTTAATAGTTTACGTTA 27

Search completed: August 25, 2003, 23:47:50  
Job time : 355.341 secs



REFERENCE AUTHORS	3 Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P., Kono, H., Akiyama, J., Nishi, K., Katsunari, T., Tashiro, H., Itoh, M., Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishino, T., Harada, A., Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K., Fujisake, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Matsubara, T., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura, S., Kawai, T., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.
TITLE	RIKEN Integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multiplexed sequencer
JOURNAL MEDLINE PUBMED REFERENCE AUTHORS	Genome Res. 10 (11), 1757-1771 (2000) 20530913 11076861 4 Kawai, J., Shinagawa, A., Shibata, K., Yoshino, M., Itoh, M., Ishii, Y., Aizawa, T., Hara, A., Fukunishi, Y., Kono, H., Adachi, J., Fukuda, S., Aizawa, K., Izawa, M., Nishi, K., Kiyosawa, H., Kondo, S., Yamana, I., Saito, T., Okazaki, Y., Gojobori, T., Bono, H., Kasukawa, T., Saito, R., Kadota, K., Matsuda, H., Ashburner, M., Batalov, S., Casavant, T., Fleischmann, W., Gaasterland, T., Glasl, C., King, B., Kochiwa, H., Kuehl, P., Lewis, S., Matsuo, Y., Nikaido, I., Pesole, G., Quackenbush, J., Schriml, L., Staudt, F., Suzuki, R., Tomita, M., Wagner, L., Washio, T., Sakai, K., Okido, T., Furuno, M., Aono, H., Balderelli, R., Barash, G., Blake, J., Boffelli, D., Bojunga, N., Carninci, P., de Bonaldo, M., Brownstein, M. J., Bull, C., Fletcher, C., Fujita, M., Gariboldi, M., Giustolich, S., Hill, D., Hofmann, M., Hume, D. A., Kamiya, M., Lee, N. H., Lyons, P., Marchionni, L., Mashima, J., Mazzarelli, J., Mombauts, P., Nordone, P., Ring, B., Ringwald, M., Rodriguez, I., Sakamoto, N., Sasaki, H., Sato, K., Schonbach, C., Seya, T., Shibata, Y., Storch, K. F., Suzuki, H., Toyooka, K., Wang, K. H., Weitz, C., Whitaker, C., Winding, L., Wynshaw-Boris, A., Yoshida, K., Hasegawa, Y., Kawai, H., Kohsaki, S. and Hayashizaki, Y.
TITLE	Functional annotation of a full-length mouse cDNA collection
JOURNAL MEDLINE PUBMED REFERENCE AUTHORS	Nature 409 (6821), 685-690 (2001) 21085660 11217851 5 The FANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team.
TITLE	Analysis of the mouse transcriptome based on functional annotation
JOURNAL PUBMED REFERENCE AUTHORS	Nature 420, 563-573 (2002) 6 (bases 1 to 2269) Adachi, J., Aizawa, K., Akimura, T., Arikawa, T., Bono, H., Carninci, P., Fukuda, S., Furuno, M., Hanagaki, T., Hara, A., Hashizume, W., Hayashida, K., Hayatsu, N., Hiramoto, K., Hirooka, T., Hirozane, T., Horii, F., Imotani, K., Ishii, Y., Itoh, M., Kigawa, I., Kasukawa, T., Kato, H., Kawai, J., Kojima, Y., Kondo, S., Kono, H., Kouda, M., Koyu, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M., Nakamura, Y., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Ohata, N., Okazaki, Y., Saito, R., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N., Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, T., Sogabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takaku-Akahira, S., Takebe, Y., Tanaka, T., Tomaru, A., Toya, T., Yasunishi, A., Muramatsu, M. and Hayashizaki, Y.
TITLE	Direct Submission
JOURNAL	Submitted (16-JUL-2001) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN genomic Sciences Center (GSC), RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail: genome-res@gs.c.riken.go.jp, URL: http://genome.gsc.riken.go.jp/, Tel: 81-45-503-9222, Fax: 81-45-503-9216)
COMMENT	cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.
FEATURES	Please visit our web site for further details. URL: http://genome.gsc.riken.go.jp/ URL: http://fantom.gsc.riken.go.jp/ Location/Qualifiers

source	1. 2269 /organism="Mus musculus" /mol_type="mRNA" /strain="C57BL/6J" /db_xref="FANTOM:DB:9430013617" /db_xref="taxon:10090" /clone="9430013617" /issue_type="embryonic body between diaphragm region and neck" /clone_1lb="RIKEN full-length enriched mouse cDNA library" /dev_stage="12 days embryo" 198, 776
CDS	/note="unnamed protein product; RAS-related G3 botulinum substrate 1 (MSD) MGI:97845, GB AK011072, evidence: BLASTN, 100%, match=987) putative" /codon_start=1 /protein_id="BAC28767.1" /db_xref="GI:26330057" /translation="MDKICVVGDAVGKTCILISYTNAPGEYIPVFDNYSNVMDGKPVNLGLMDTAQEDYDRLPLSPQTDVFLCFSLVSPASENVRKVPYRHHCPNPPIILVGTIKLDRDKDITKLEKILTPITYPOGLMAKKEIGAVKYLECSALTORGLTVEDEAIRAVICPPPVKRRKRCILL" 2252, 2257 /note="putative" 2269
polyA_site	polyA_site 2269
BASE COUNT	555 a 569 c 502 g 643 t
ORIGIN	
Query Match	67.3%; Score 851.6; DB 11; Length 2269; Best Local Similarity 85.7%; Pred. No. 1.1e-102; Matches 1081; Conservative 0; Mismatches 129; Indels 52; Gaps 10;
QY	50 ATGATTCACAGATTTTATTAAGTCATCATCATCAAAACACTACTGCTAGTACGAAAA 109 
Db	2269 ATGATTCACAGATTTTATTAAGTCATCATCATCAAAACACTACTGCTAGTACGAAAA 2210 
QY	110 GATCATATGTAACAACTCCACAAATTCGCACTGCAATGTAACAAAGTTTGTCTACT 169 
Db	2209 GATCATATGTAACAACTCCACAAATTCGCACTGCAATGTAACAAAGTTTGTCTACT 2150 
QY	170 GGTGCAAGGCGCCACACGCTGTCTTCCAGAGAGTAGTGTGACAGAGCGCTTGG 229 
Db	2149 GGTGCAAGGCGCCACACGCTGTCTTCCAGAGAGTAGTGTGACAGAGCGCTTGG 2090 
QY	230 CACTAGCGCTTGACAGAACCTCAGACCCCAAG-----TACC 268 
Db	2089 CACTAGCAGTTGACAGAACCTCAGACCCCAAGAACATCTCTAGAGAGAGCATGACA 2030 
QY	269 GGAAGCATGTCTCCGCTGGGTGAGGTCTAGAGGGGGCGCATCAATCAGATCAGTGT 328 
Db	2029 GGAAGCATGTCTCCGCTGGGTGAGGTCTAGAGGGGGCGCATCAATCAGATCAGTGT 1973 
QY	329 TGTACTCTGGCAGACAGATGATGTTTCAATATATCAATTAATTAAGTGAAG 388 
Db	1972 TGTACTCTGGTAAAGGTTTCAATATATCAATTAATTAAGTGAAG 1913 
QY	389 CCGCAGACGCTGATTTCTACACCCAGTTACTAGAAAAAGGAAGGAGTACTAGT-CAGCT 447 
Db	1912 CTGCAACACATGATTTGACACCTAGTGTGTAAGAACTAAGAGACATTAAGTACTGT 1853 
QY	448 GAGTAAAGAGGTGAAGACAGACGACGCTTACTATCTACCAAAAAAATCTCGAA- 506 
Db	1852 GAGTAAAGGTGAAGACAGACGACGCTTACTATCTACCAAAAAAATCTCGAA- 1793 
QY	507 ----TGATATATAGAAAGATCTTATAGTACAGTACAGATATTTGCTTAAAGG 562 
Db	1792 TCTCTGCTTATGAGAAAGATCTTATAGTACAGTACAGATATTTGCTTAAAGG-GG 1734 
QY	563 GGTCTTAAAGAAAGCACTGTGAAGTTAGCACTGTGAGAGATGCGCAGTTTAAATATG 622 
Db	1733 GGTCTTAAAGAAAGCACTGTGAAGTTAG-MACTGTGAAGGTGGCCAGTTTAAATATG 1675 

```

OY 623 ACTCAAGCCCATCTGGGAGGAGCAGCAGGGGGAGGGGCTCAG----- 671
DB 1674 ACTCAAGCCCATCTGGGAGGAGCAGCAGGGGGAGGGGCTCAG----- 1615
OY 672 -AGAGACTGATTAAGATGCGCCATTTGTCACTCTGTTTACAGAAATTAACGTTAA 730
DB 1614 AAGAGACTGATTAAGATGCGCCATTTGTCACTCTGTTTACAGAAATTAACGTTAA 1555
OY 731 AAGGCTTACCGGTGACACTTTTATTCAGTTGAATTAATTCAGTAATTAAGTTAA 790
DB 1554 AAGGCTTACCGGTGACACTTTTATTCAGTTGAATTAATTCAGTAATTAAGTTAA 1495
OY 791 TTAATCTCTACTTCAATTAATTAAGTCAAAATTAATCTCTCTCTCTGTTGATGACGCTGTTTC 850
DB 1494 GTAATCTCTACTTCAATTAATTAAGTCAAAATTAATCTCTCTCTCTGTTGATGACGCTGTTTC - 1436
OY 851 ACACACTGACCCAGCAGCAGCCACAGCAGTGGAGCAAGTAATCTCTGTTAGGCAACACAG 910
DB 1435 -CACACTGACCCAGCAGCAGCCACAGCAGTGGAGCAAGTAATCTCTGTTAGGCAACACAG 1377
OY 911 GAGCAGAGTTCTGTTCAAGCCTGAGAGCGGTCAGCTGTTATTTAGAACCTCAG 970
DB 1376 GAGCAGAGTTCTGTTCAAGCCTGAGAGCGGTCAGCTGTTATTTAGAACCTCAG 1317
OY 971 TATGAATTCAGAGCAGCAGCTGTTACACCCATC---GTGACGTACAGTACAAAGTTACG 1027
DB 1316 TATGAATTCAGAGCAGCAGCTGTTACACCCATC---GTGACGTACAGTACAAAGTTACG 1257
OY 1028 TATATGACATGGCGGTGATTAAGTACAGGTGCTTACATGACGCTGATTAAGAGGC 1087
DB 1256 TATATGACATGGCGGTGATTAAGTACAGGTGCTTACATGACGCTGATTAAGAGGC 1197
OY 1088 TGTGCTGTGTCACAGCGTGTGGAGGAGTACGGGAGGTCGACCCCTGAGCCAGAGCT 1147
DB 1136 TGTGCTGTGTCACAGCGTGTGGAGGAGTACGGGAGGTCGACCCCTGAGCCAGAGCT 1137
OY 1148 GCACTCTCTTAAAGACAAA---GTCTCTCAACAGCTAGTCTTACGTTCTCAGCAG 1204
DB 1136 GCGGCTCTCTTAAAGACAAA---GTCTCTCAACAGCTAGTCTTACGTTCTCAGCAGC 1077
OY 1205 AAGCAACTTACTTCAAGAGTATTTGGCAATTTCTTATCTGACGACAGAAAGGGATT 1264
DB 1076 AAGCAACTTACTTCAAGAGTATTTGGCAATTTCTTATCTGACGACAGAAAGGGATT 1017
OY 1265 TT 1266
DB 1016 TT 1015

```

```

RESULT 2
LOCUS AK081613/c
DEFINITION Mus musculus 16 days embryo head cDNA, RIKEN full-length enriched
sublibrary 1, full insert sequence.
ACCESSION AK081613
VERSION AK081613.1 GI:26349264
KEYWORDS HTC; CAP trapper.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sclerognathi; Muridae; Murinae; Mus.
REFERENCE
AUTHORS Carninci, P., and Hayashizaki, Y.
TITLE High-efficiency full-length cDNA cloning
JOURNAL Meth. Enzymol. 303, 19-44 (1999)
MEDLINE 99279253
PUBMED 10349636
REFERENCE
AUTHORS Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,
Itoh, M., Kono, H., Okazaki, Y., Muramatsu, M., and Hayashizaki, Y.
TITLE Normalization and subtraction of cap-trapper-selected cDNAs to

```

JOURNAL  
MEDLINE  
PUBMED  
REFERENCE  
AUTHORS

prepare full-length cDNA libraries for rapid discovery of new genes  
Genome Res. 10 (10), 1617-1630 (2000)  
2049374  
11042155

3 Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P.,  
Kono, H., Akiyama, J., Nishi, K., Katsunai, T., Tashiro, H., Itoh, M.,  
Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishino, T., Harada, A.,  
Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kasahara, K.,  
Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watanabe, M.,  
Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura, S., Kawai, J.,  
Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A., and Hayashizaki, Y.  
RIKEN integrated sequence analysis (RISA) system -384-format  
sequencing pipeline with 384 multicapillary sequencer  
Genome Res. 10 (11), 1757-1771 (2000)  
20530913  
11076861

JOURNAL  
MEDLINE  
PUBMED  
REFERENCE  
AUTHORS

4 Kawai, J., Shinagawa, A., Shibata, K., Yoshino, M., Itoh, M., Ishii, Y.,  
Arakawa, T., Hara, A., Fukunishi, Y., Kono, H., Adachi, J., Fukuda, S.,  
Aizawa, K., Izawa, K., Nishi, K., Kiyosawa, H., Kondo, S., Yamane, I.,  
Saito, T., Okazaki, Y., Gojobori, T., Bono, H., Kasukawa, T., Saito, R.,  
Kadota, K., Matsuda, H., Ashburner, M., Batalov, S., Casavant, T.,  
Fleischmann, W., Gaasterland, T., Gissi, C., King, B., Kochiya, H.,  
Kuehl, P., Lewis, S., Matsuo, Y., Nikaido, I., Pesole, G.,  
Quackenbush, J., Schriml, L. M., Staudt, F., Suzuki, R., Tomita, M.,  
Wagner, L., Washio, T., Sakai, K., Okido, T., Furuno, M., Aono, H.,  
Baldarelli, R., Barsh, G., Blake, J., Boffelli, D., Bojunga, N.,  
Carninci, P., de Bonaldo, M. F., Brownstein, M. J., But, C.,  
Fletcher, C., Fujita, M., Gariboldi, M., Gustincich, S., Hill, D.,  
Hofmann, M., Hume, D. A., Kamiya, M., Lee, N. H., Lyons, P.,  
Marchionni, L., Mashima, J., Mazzerelli, J. J., Kombearts, P., Nordone, P.,  
Rind, B., Ringwald, C., Seya, T., Shibata, Y., Storch, K. F., Suzuki, H.,  
Sato, K., Schonbach, C., Seya, T., Shibata, Y., Storch, K. F., Suzuki, H.,  
Toyo-Oka, K., Wang, K. H., Welte, C., Whitlaker, C., Wilming, L.,  
Wynshaw-Boris, A., Yoshida, K., Hasegawa, Y., Kawai, H., Kohlsuk, S.,  
and Hayashizaki, Y.  
Functional annotation of a full-length mouse cDNA collection  
Nature 409 (6821), 685-690 (2001)  
21085660  
11217851

JOURNAL  
MEDLINE  
PUBMED  
REFERENCE  
AUTHORS

5 The FANTOM Consortium and the RIKEN Genome Exploration Research  
Group Phase I & II Team.  
Analysis of the mouse transcriptome based on functional annotation  
of 60,770 full-length cDNAs  
Nature 420, 563-573 (2002)

JOURNAL  
MEDLINE  
PUBMED  
REFERENCE  
AUTHORS

6 (bases 1 to 1796)  
Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Bono, H., Carninci, P.,  
Fukuda, S., Furuno, M., Hanagaki, T., Hara, A., Hashizume, M.,  
Hayashida, K., Hayatsu, N., Hiramoto, K., Hiraoka, T., Hirozane, T.,  
Hori, F., Imoto, K., Ishii, Y., Itoh, M., Kagawa, I., Kasukawa, T.,  
Kato, H., Kawai, J., Kojima, Y., Kondo, S., Kono, H., Kouda, M.,  
Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M.,  
Nakamura, M., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Ohsato, N.,  
Okazaki, Y., Saito, R., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N.,  
Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, T.,  
Sogabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takaku-Akihira, S.,  
Takeda, Y., Tanaka, T., Tomaru, A., Toya, T., Yasunishi, A.,  
Muramatsu, M., and Hayashizaki, Y.  
Direct Subdivision

JOURNAL

Submitted (16-Apr-2002) Yoshihide Hayashizaki, The Institute of  
Physical and Chemical Research (RIKEN), Laboratory for Genome  
Exploration Research Group, RIKEN Genomic Sciences Center (GSC),  
RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,  
Kanagawa 230-0045, Japan (E-mail: genome-res@gsr.riken.go.jp,  
URL: http://genome.gsc.riken.go.jp/, Tel: 81-45-503-9222,  
Fax: 81-45-503-9216)  
cDNA library was prepared and sequenced in Mouse Genome  
Encyclopedia Project of Genome Exploration Research Group in RIKEN  
Genomic Sciences Center and Genome Science Laboratory in RIKEN.  
Division of Experimental Animal Research in Riken contributed to  
prepare mouse tissues.

Please visit our web site for further details.

URL: <http://genome.gsc.riken.go.jp/>  
URL: <http://fantom.gsc.riken.go.jp/>

## FEATURES

## Source

1. 1796

Location/Qualifiers

/organism="Mus musculus"

/mol\_type="mRNA"

/strain="C57BL/6J"

/db\_xref="FANTOM\_DB:C130051H24"

/db\_xref="taxon:10090"

/clone="C130051H24"

/issue\_type="head"

/clone\_lib="Riken full-length enriched mouse cDNA library"

/dev\_stage="16 days embryo"

<1.304

/note="unnamed protein product; RAS-related C3 botulinum

substrate 1 (MGI:MGI:97845, GB|AK01072, evidence: BLASTN,

100%, match=987)

putative"

/protein\_id="BAC38272.1"

/db\_xref="GI:26349265"

/translation="VRKMYPEVRHHCNPPIIVGKLDLRDQKDIIEKLEKLLP

ITVPGIAMAKEIGAVKYLECSALTQRLKTFDEAIFRAVLCPPYKRRKCLLL"

1781..1786

/note="putative"

polyA\_site

1796

/note="putative"

BASE COUNT

470 a 413 c 371 g 542 t

ORIGIN

Query Match 67.2%; Score 850.2; DB 11; Length 1796;

Best Local Similarity 85.6%; Pred. No. 2e-102;

Matches 1080; Conservative 0; Mismatches 128; Indels 53; Gaps 10;

```

  52 GATTCAAGATTATTAAGTCATACATGCAAAACATCTCTAATCTGATAGCAAAAGA 111
      |||
  1796 GATTCAGAGATTATTAATTCATCATGCAAAACATCTCTAATCTGATAGCAAAAGA 1737
      |||
  112 TCATGTAAGAAACACCTCCAAATTCGCAACTGTCATTTGAAAGATTGTTCTAGTGG 171
      |||
  1736 TCATGTAAGAAACACCTCCAAATTCGCAACTGTCATTTGAAAGATTGTTCTAGTGG 1677
      |||
  172 TCGAAGAGCCCAACACTGTCTTTCGAGTGAAGTTGTTGTCAGAGAGGGGTTAGCA 231
      |||
  1676 TTGAAGAGCCCAACACTGTCTTTCGAGTGAAGTTGTTGTCAGAGAGGGGTTAGCA 1617
      |||
  232 CTAGCGCTTGACAGAACCTCAGAACCCAAAG-----TACCGG 270
      |||
  1616 CTAGCAGTTGACAGAACCTCAGAACCCCAAGAACATCTCTAGGAGAGCATGACAGG 1557
      |||
  271 AAGCATCTGTCGGCGTGGGTGAGGTCCTAGAGGGGGCGGCATCATCATGACAGTGTG 330
      |||
  1556 AAGCGTCTGTCAGAGTGGGAGGCTTGA---GGCAGCATTTAGTCTCATGACAGTGTG 1500
      |||
  331 GTAATGCGGAAGCACTGATGTTTCAAGATATCTAAATAGTTAAAACTGTAAAGCC 390
      |||
  1499 GTAATGCGGAAGCACTGATGTTTCAAGATATCTAAATAGTTAAAACTGTAAAGCC 1440
      |||
  391 GCAAGCAGGTATTTCTACACCCAGTACTAGAAAAAGAGGAGACACTAGT-CAGTGA 449
      |||
  1439 GCAAGCAGGTATTTCTACACCCAGTACTAGAAAAAGAGGAGACACTAGTACTCTGTA 1380
      |||
  450 GTAAGGAGGTGAAAAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 506
      |||
  1379 GTAAGGAGGTGAAAAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1320
      |||
  507 ---TGCAATTATCAAGAAAGATCTTATAGTACAGTCAAGACATATTTGCTGTAAGAGGG 563
      |||
  1319 CTGCGGCTTATCAAGAAAGATCTTATAGTACAGTCAAGACATATTTGCTGTAAGAA-GGG 1261
      |||
  564 GTCCATAAAGAAAGACCTTGCTAAGTTCAGAACTGTGAGGATGGCGATTAAATATGGA 623
      |||
```

```

Db 1260 GTTCTAAAGAAAGACACTTGTCTAAGTTAG-ACCTGAGAGGTGGCCAGTTTAATATGCA 1202
      |||
  624 CTCAGCGCCCATCTGGGGAGGAGACAGAGGGGAAAGGGGGCTCAAG----- 671
      |||
  1201 CTCAGCGCCCATCTGGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1142
      |||
  672 AGAGCAGCTATTAAGATTCGGCCATTGTCATCTACGTGTTGACAGAAATTAACCGTTAA 731
      |||
  1141 AGAGCAGCTATTAAGATTCGGCCATTGTCATCTACGTGTTGACAGAAATTAACCGTTAA 1082
      |||
  732 AACCTTACCCGTCAGACACTTTTATTCAGTTGAATTTCTCCATGTACAAATGTAGTGAAT 791
      |||
  1081 AACCTTACCCGTCAGACACTTTTATTCAGTTGAATTTCTCCATGTACAAATGTAGTGAAT 1022
      |||
  792 TAACTCTACTATATTAATTAAGTCAAAATTAATCTGTCCTGCTTGTAGTACGTCGTTTCA 851
      |||
  1021 TAACTCTACTATATTAATTAAGTCAAAATTAATCTGTCCTGCTTGTAGTACATCTGTTT -- 964
      |||
  852 CACAGCTCACCCAGCAGCAGCCAGCAGTACAGAAAGAAATTTCTGTTAGAGGACAGAG 911
      |||
  963 CACAGCTCACCCAGCAGCAGCCAGCAGTACAGAAAGAAATTTCTGTTAGAGGACAGAG 904
      |||
  912 AGCCAGAGTTCTGTTCAAAAGCCTGAGAAAGCCGTCAGAGTGTGATTTTATAGAACTACT 971
      |||
  903 AACGAGAGTTCTGTTCAAAAGCCTGAGAAAGCCTGAGAGTGTGATTTTATAGAACTACT 844
      |||
  972 ATGAATCAAGAGCAGAGAGCTGTTACACCATC---GTGACGTACAGTACAAATTAAGT 1028
      |||
  843 ATGAATCAAGAGCAGAGAGCTGTTACACCATC---GTGACGTACAGTACAAATTAAGT 784
      |||
  1029 ATGAGCATGAGGCTGATTAATTAAGTACAGTGTGATTAAGTGTGATTAAGTGTGATTAAG 1088
      |||
  783 ATGAGCATGAGGCTGATTAATTAAGTACAGTGTGATTAAGTGTGATTAAGTGTGATTAAG 724
      |||
  1089 GTGCTGTGTCACAGAGTCTGGAGCTTACGAGAGGCTTGCAGCCCTGAGAGGAGAGCTG 1148
      |||
  723 GTGCTGTGTCACAGAGTCTGGAGCTTACGAGAGGCTTGCAGCCCTGAGAGGAGAGCTG 664
      |||
  1149 CAGCTCTTCTTAAGGACAA--GTCTCTCAACAGCTTAAGTGTGATTAAGTGTGATTAAG 1205
      |||
  663 CAGCTCTTCTTAAGGACAA--GTCTCTCAACAGCTTAAGTGTGATTAAGTGTGATTAAG 604
      |||
  1206 ACGCAACTTGAATTCACAAAGTATTTGGCAATCTTAATCTGAGCAAAATTAAGGGATT 1265
      |||
  603 ACGCAACTTGAATTCACAAAGTATTTGGCAATCTTAATCTGAGCAAAATTAAGGGATT 544
      |||
  1266 T 1266
      |||
  543 T 543
      |||
  Db 543 T 543
      |||

RESULT 3
AK047969/c 2270 bp mRNA linear HRC 05-DEC-2002
LOCUS AK047969
DEFINITION Mus musculus 16 days embryo head cDNA, RIKEN full-length enriched
library, clone: C130025F01 product: RAS-related C3 botulinum
substrate 1, full insert sequence.
ACCESSION AK047969
VERSION AK047969.1 GI:26339063
KEYWORDS HRC; CAP trapper.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1
AUTHORS Carninci, P. and Hayashizaki, Y.
TITLE High-efficiency full-length cDNA cloning
JOURNAL Meth. Enzymol. 303, 19-44 (1999)
MEDLINE 99279253
PubMed 10349636
REFERENCE 2
AUTHORS Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,
Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
```

TITLE	JOURNAL	MEDLINE	PUBMED	REFERENCE	AUTHORS
Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes	Genome Res.	10 (10), 1617-1630 (2000)	11042155		Shibata, K., Itoh, M., Alzawa, K., Nagaoka, S., Sasaki, N., Carolin, P., Sumi, N., Aliyama, J., Nishi, K., Kitasuna, T., Tashiro, H., Itoh, M., Kuni, H., Ishii, Y., Nakamura, S., Hazama, M., Nishino, T., Harada, A., Yamamoto, R., Matsunoto, H., Sekaguchi, S., Ikegami, T., Kashiwagi, K., Fujikawa, S., Inoue, K., Togawa, Y., Iwawa, M., Ohara, E., Matsubiki, K., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsunaga, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.
RIKEN Integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer	Genome Res.	10 (11), 1757-1771 (2000)			
	JOURNAL	MEDLINE	PUBMED	REFERENCE	AUTHORS
	20530913				
	11076861				
	4				
	Kawai, J., Shinagawa, A., Shibata, K., Yoshino, M., Itoh, M., Ishii, Y.,				



QY	565	TCCTAAAGAAAGACATTGCTCAACTTGTGGCAACCTGTGGAGGAATGGCAGTTTAAATATGAC	624
Db	1733	TTCTAAAGAAAGACACTTGTCTAAAGTAC- AACTGTGAGGTGGCCAGTTTAAATATGAC	1675
QY	625	TCAACGCCCATCTGTGGGAGGGAGACAGCAGGGGGAGGGGGCTCAAG-----A	672
Db	1674	TCAACGCCCATCTGTGGGAGGGAGCGCAGTGTAGAGGGTGGGGGGCGCGAAGGATGTCTCA	1615
QY	733	AGCTTTACCCCTGACACTTTTATTCAGTTGGAATTACTCCATGTCAATATGTATTAATT	792
Db	1554	AGCTTTACCCCTGACACTTTTATTCAGTTGGAATTACTCCATGTCAATATGTATTAAGT	1495
QY	793	AATCTCACTTCATATTAATAGTCAAAAATACTGTCTGTCTCTCTTTGATGACGTGTCTTAC	852
Db	1494	AATCTCACTTCATATTAATAGTCAAAAATACTGTCTGTCTCTCTTTGATCATGTGTGTCTT--C	1437
QY	853	ACACTGCACCCGAGACCCCGACGACTGAGGACAGAAATCACTCGTTAGAGGACACAGGA	912
Db	1436	ACACTGCACCCGAGACCCCGACGACTGAGGACAGAAATCACTCATTAAGGACACACAGGA	1377
QY	913	GCCAGAGTTCGTCTCAAAAGCCCTGCAGAACCCGCTGACGTGTATTTTGGAGACCTACTA	972
Db	1376	ACGAGAGTTCGTCTCAAAAGCTGTGCAAAAAGCTAGTCGGCTGTCTTTTGGAAACCTACTA	1317
QY	973	TGAAATCAAAAGCAGACAGCTTTACACCATC---GTGACGTACAGTCAAAAGTTACGTA	1029
Db	1316	TGAAATCAAAAGCTGAGCTGTCACTACCTACACTGTGACGTACAGTCAAAAGTTACTTA	1257
QY	1030	ATGACATAGGCGCTCATTAAGTTACAGGTCGCTTACATGGCAGCAGTGTCTTAAAGAGCGTG	1089
Db	1256	ATGAAACATGGCGCTCATTAAGTTACAGGTCGCTTACATGGCAGCAGTGTCTTAAAGAGCGTG	1197
QY	1090	TGCTGTGTCAACAGCGTCTGGGAGCTACGGAGGGTCTGTGACCCCTGAGCCCGAAGCTGC	1149
Db	1196	TGCTGTGTCAACAGCGTCTGGGAGCTACGGAGGGTCTGTGACCCCTGAGCCCGAAGCTGC	1137
QY	1150	AGCTTCTCTTAAGACAAA---GTCTCTCAACAGCTTAAGTGTCTTACGTCTTCTACACACAA	1206
Db	1136	GGCTTCTCTTAAGACAAAGAGTCTCTCTCAAGAGATTATAGTGTGTGTCTTCTACACACAA	1077
QY	1207	CGCAACTAGTCTCAAGAGATTATTTGGCAATCTTAAATCTGAGCAAGAAATAGGGGATTTT	1266
Db	1076	CGCAACTAGTCTCAAGAGATTATTTGGCAATTTTAAATCTGAGCAAGAAATAGGGGATTTT	1017
RESULT 4			
AK076023/c			
LOCUS	AK076023	2308 bp	mRNA
DEFINITION	Mus musculus 10 days embryo whole body cDNA, RIKEN full-length enriched library, clone:2610100016 product:RAS-related C3 botulinum substrate 1, full insert sequence.		
ACCESSION	AK076023		
VERSION	AK076023.1	GI:26344957	
KEYWORDS	HTC; CAP trapper.		
SOURCE	Mus musculus (house mouse)		
ORGANISM	Mus musculus		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.		
AUTHORS	Carninci, P. and Hayashizaki, Y.		
TITLE	High-efficiency full-length cDNA cloning		
JOURNAL	Meth. Enzymol. 303, 19-44 (1999)		
MEDLINE	99279253		
PUBMED	10349636		
REFERENCE	2		
AUTHORS	Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Kono, H., Okeazaki, Y., Muramatsu, M. and Hayashizaki, Y.		
TITLE	Normalization and subtraction of cap-trapper-selected cDNAs to		

JOURNAL MEDLINE PUBMED REFERENCE AUTHORS	Prepare full-length cDNA libraries for rapid discovery of new genes Genome Res. 10 (10), 1617-1630 (2000) 20499374 11042159
TITLE	3
JOURNAL MEDLINE PUBMED REFERENCE AUTHORS	Shibata,K., Itoh,M., Aizawa,K., Nagaoaka,S., Sasaki,N., Carninci,P., Konno,H., Akiyama,J., Nishi,K., Kitagawa,T., Tashiro,H., Itoh,M., Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishino,T., Harada,S., Yamanoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K., Fujiwara,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Matshiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsura,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kita,A. and Hayashizaki,Y. RIKEN Integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multichipillary sequencer Genome Res. 10 (11), 1757-1771 (2000) 20530913 11076861
TITLE	4
JOURNAL MEDLINE PUBMED REFERENCE AUTHORS	Kawai,J., Shingawa,A., Shibata,K., Yoshino,M., Itoh,M., Ishii,Y., Aizawa,T., Hara,A., Fukunishi,Y., Konno,H., Adachi,J., Fukuda,S., Aizawa,K., Izawa,M., Nishi,K., Kiyosawa,H., Kondo,S., Yamanaka,I., Saito,T., Okazaki,Y., Gotohori,T., Bono,H., Katsukawa,T., Saito,R., Kodori,K., Matsuda,H., Ashburner,M., Batalov,S., Casavant,T., Fleischmann,M., Gaasterland,T., Gissi,C., King,B., Kochiya,H., Kuehl,P., Lewis,S., Matsuo,Y., Nakado,T., Pezole,G., Quackenbush,J., Schriml,L.M., Staudt,F., Suzuki,R., Tomita,M., Wagner,L., Washio,T., Sakai,K., Okido,T., Furuno,M., Aono,H., Baldarelli,R., Barish,G., Blake,J., Boffelli,D., Bojunga,N., Carninci,P., de Bonaldo,M.F., Brownstein,M.J., But,C., Fletcher,C., Fujita,M., Gariboldi,M., Gustincich,S., Hill,D., Hochmann,L., Hume,D.A., Kamlay,M., Lee,N.H., Lyons,P., Marchionni,L., Mashima,J., Mazzarelli,J., Mombaerts,P., Nordone,P., Rling,B., Ringwald,M., Rodriguez,I., Sakamoto,N., Sasaki,H., Sato,K., Schonbach,C., Seya,T., Shibata,Y., Storch,K.F., Suzuki,H., Toyo-Oka,K., Wang,K.H., Weitz,C., Willeker,C., Wilming,L., Wyshaw-Bois,A., Yoshida,K., Hasegawa,Y., Kawaji,H., Kohsaki,S. and Hayashizaki,Y. Functional annotation of a full-length mouse cDNA collection Nature 409 (6821), 685-690 (2001) 21085660 11217851
TITLE	5
JOURNAL MEDLINE PUBMED REFERENCE AUTHORS	The FANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team. Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs Nature 420, 563-573 (2002) 6 (bases 1 to 2308)
JOURNAL REFERENCE AUTHORS	Adachi,J., Aizawa,K., Akahira,S., Akimura,T., Aono,H., Arai,A., Araiawa,T., Bono,H., Carninci,P., Fukuda,S., Fukunishi,Y., Furuno,M., Hanagaki,T., Hara,A., Hayatsu,N., Hiramoto,K., Hiraoaka,T., Hori,F., Imotani,K., Ishii,Y., Itoh,M., Izawa,M., Kasukawa,T., Kato,H., Kawai,J., Kojima,Y., Konno,H., Kouda,M., Koya,S., Kurihara,C., Matsuyama,T., Miyazaki,A., Nishi,K., Nomura,K., Numazaki,R., Ono,M., Okazaki,Y., Okido,T., Owa,C., Saito,H., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki,D., Shibata,K., Shibata,Y., Shingawa,A., Shiraki,T., Sogabe,Y., Suzuki,H., Tagami,M., Tagawa,A., Takeuchi,F., Tanaka,T., Tejima,Y., Toyo,T., Yamamura,T., Yamanaka,T., Yasunishi,A., Yoshida,K., Yoshino,M., Muramatsu,M. and Hayashizaki,Y. Direct Submission Submitted (16-APR-2002) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail:genome-res@gscc.riken.go.jp, URL:http://genome.gsc.riken.go.jp/, Tel:81-45-503-9222, Fax:81-45-503-9216)
COMMENT	cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in RIKEN Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.



MEDLINE 97044477  
 PUBMED 8889548  
 COMMENT Contact: Soares, MB  
 Coordinated Laboratory for Computational Genomics  
 University of Iowa  
 375 Newton Road, 4156 MEBRF, Iowa City, IA 52242, USA  
 Tel: 319 335 8250  
 Fax: 319 335 9565  
 Email: bento-soares@uiowa.edu  
 The sequence contained an oligo-dT track that was present in the oligonucleotide that was used to prime the synthesis of first strand cDNA and therefore this may represent a bonafide poly A tail. The sequence tag present in the cDNA between the NotI site and the oligo-dT track served to verify it as a clone from the normalized osteoblast library cDNA library Preparation: M.B. Soares Lab Clone distribution: clones will be available through Research Genetics (www.resgen.com)  
 Seq primer: M13 Forward  
 POLYA=yes.

FEATURES  
 Source Location/Qualifiers  
 1..760  
 /organism="Rattus norvegicus"  
 /mol\_type="mRNA"  
 /strain="Sprague-Dawley"  
 /db\_xref="taxon:10116"  
 /clone="UI-R-DRI-CKY-h-11-0-UI"  
 /dev\_stage="adult"  
 /lab\_host="DH10B (Life Technologies)"  
 /clone\_lib="UI-R-DRI"  
 /note="Vector: pT73D-Pac (Pharmacia) with a modified polylinker. Site 1: Not I; Site 2: Eco RI; The UI-R-DRI library is a normalized Rat osteoblast library (nreo) constructed in pT37 vector according to the procedure described by Bonaldo, Lennon and Soares (Normalization and Subtraction: Two Approaches to Facilitate Gene Discovery. Genome Research 6: 791-806, 1996). The oligonucleotide used to prime first strand synthesis contained the sequence tag AAGATATCAA between the Not I cloning site and dT18 stretch. The Rat Osteoblast tissue was provided by Lien & Stein of the University of Massachusetts Medical School.  
 TAG\_LIB=UI-R-DRI  
 TAG\_TISSUE=osteoblast  
 TAG\_SEQ=AAGATATCAA"

BASE COUNT 244 a 143 c 173 g 198 t 2 others

ORIGIN

Query Match 58.2%; Score 736.4; DB 13; Length 760;  
 Best Local Similarity 98.7%; Pred. No. 2.4e-87;  
 Matches 752; Conservative 0; Mismatches 8; Indels 2; Gaps 1;

34 TTTTCTTTTCTTTTATGATTCAGGATTTATTAGTCATACATGCAAAACATAGTCT 93  
 |||||||  
 1 TTTTCTTTTCTTTTATGATTCAGGATTTATTAGTCATACATGCAAAACATAGTCT 60  
 |||||||

94 AACGCGATTGCAAAAGATCAATGTAATAAACACGCCCAACTGCTTGGCAAGTCAATGAA 153  
 |||||||  
 61 AACGCGATTGCAAAAGATCAATGTAATAAACACGCCCAACTGCTTGGCAAGTCAATGAA 120  
 |||||||

154 AAAAGTTTCTTCTAGTGTGCAAAAGGCCCAACTGCTTGGCAAGTCAATGAA 213  
 |||||||  
 121 AAAAGTTTCTTCTAGTGTGCAAAAGGCCCAACTGCTTGGCAAGTCAATGAA 180  
 |||||||

214 TACAGAAAGCGGTTAGACACTAGCGCTTGACAGAACTCAGACAGCCCAAGGTACGGGAG 273  
 |||||||  
 181 TACAGAAAGCGGTTAGACACTAGCGCTTGACAGAACTCAGACAGCCCAAGGTACGGGAG 240  
 |||||||

274 CATGTGTCGGGTTGAGGTCTAGAGGGGGCGGCATCATCATGAGAGAGTTGGTA 333  
 |||||||  
 241 CATGTGTCGGGTTGAGGTCTAGAGGGGGCGGCATCATCATGAGAGAGTTGGTA 300  
 |||||||

334 CTCTGGCAAGACAGTGTCTTGAATATCTAAATAGTTTAAAACTGTAAAGCCGCA 393  
 |||||||

Db 301 CTCTGGCAAGACAGTGTCTTGAATATCTAAATAGTTTAAAACTGTACAGCCGCA 360

Qy 394 GCACGTGATTTCTACACCCAGTTACTAGAAAAGAAAGGAGCACTAGTCACTGATTA 453  
 |||||||

Db 361 GCACGTGATTTCTACACCCAGTTACTAGAAAAGAAAGGAGCACTAGTCACTGATTA 420  
 |||||||

Qy 454 AGAAGGTGAAAACAGAAAGCACTTCTACTATCTACCAAAAAATCTCGAATGCATTA 513  
 |||||||

Db 421 AGAAGGTGAAAACAGAAAGCACTTCTACTATCTACCAAAAAATCTCGAATGCATTA 480  
 |||||||

Qy 514 TCAGAAAGATCTTATGATCAGGTCAAGATATTTGCTTGAAGAGGGGCTTAAGA 573  
 |||||||

Db 481 TCAGAAAGATCTTATGATCAGGTCAAGATATTTGCTTGAAGAGGGGCTTAAAGA 540  
 |||||||

Qy 574 AAAGCACTTCTTAAGTTAGCACTGTGAGAGTGGCCAGTTAAATATGCACTCAAGCC 633  
 |||||||

Db 541 AAAGCACTTCTTAAGTTAGCACTGTGAGAGTGGCCAGTTAAATATGCACTCAAGCC 600  
 |||||||

Qy 634 CATCTGGGAGGAGCAGACAGAGGGGAGGGGCTCAAGAGACACTGATTAATGCGCC 693  
 |||||||

Db 601 CATCTGGGAGGAGCAGACAGAGGGGAGGGGCTCAAGAGACACTGATTAATGCGCC 660  
 |||||||

Qy 694 ATTGTCATCTACTGTTTGCAGAAATTAACGTTAAAGCTTACCGGTGACACTTT 753  
 |||||||

Db 661 ATTGTCATCTACTGTTTGCAGAAATTA--CCGTTAAAGCTTACCGGTGACACTTT 718  
 |||||||

Qy 754 ATTCACTGAATTAATCTCATGTAATGATGATTAATTAAT 795  
 |||||||

Db 719 ATTCACTGAATTAATCTCATGTAATGATGATTAATTAAT 760  
 |||||||

RESULT 6  
 BM389059  
 LOCUS 687 bp mRNA linear EST 17-JAN-2002  
 DEFINITION UI-R-D20-CKO-j-07-0-UI.s1 UI-R-D20 Rattus norvegicus cDNA clone  
 BM389059  
 ACCESSION UI-R-D20-CKO-j-07-0-UI 3', mRNA sequence.  
 VERSION BM389059.1 GI:18189112  
 KEYWORDS EST.  
 SOURCE Rattus norvegicus (Norway rat)  
 ORGANISM Rattus norvegicus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Murinae; Rattus.

REFERENCE 1 (bases 1 to 687)  
 AUTHORS Bonaldo,M.F., Lennon,G. and Soares,M.B.  
 TITLE Normalization and subtraction: two approaches to facilitate gene discovery  
 JOURNAL Genome Res. 6 (9), 791-806 (1996)  
 MEDLINE 97044477  
 PUBMED 8889548

COMMENT Contact: Soares, MB  
 Coordinated Laboratory for Computational Genomics  
 University of Iowa  
 375 Newton Road, 4156 MEBRF, Iowa City, IA 52242, USA  
 Tel: 319 335 8250  
 Fax: 319 335 9565  
 Email: bento-soares@uiowa.edu  
 The sequence contained an oligo-dT track that was present in the oligonucleotide that was used to prime the synthesis of first strand cDNA and therefore this may represent a bonafide poly A tail. The sequence tag present in the cDNA between the NotI site and the oligo-dT track served to verify it as a clone from the non-normalized cartilaginous tumor library cDNA library Preparation: M.B. Soares Lab Clone distribution: clones will be available through Research Genetics (www.resgen.com)  
 Seq primer: M13 Forward  
 POLYA=yes.

FEATURES  
 Source Location/Qualifiers  
 1..687  
 /organism="Rattus norvegicus"  
 /mol\_type="mRNA"  
 /strain="Sprague-Dawley"



```

Db      62  TAACTCATTAGGAAAAAGATCAATGTAACAACTCAGCAATCTGCAACTGTCATTTGA 121
QY      153  AAAAAGTTTGTCTAGTGTGCGAAAAGGCCAACACTGTGTTCTTGCCAGTAGTTAGTT 212
Db      122  AAAAATTGTTCTTACTGGTCGAAAGGCCCAACACTGTGTTCTTGCCAGTAGTTAGTT 181
QY      213  GTACAGAACGGCGTTAGACCTAGCGCTTGACAGAACCTCAGACCCAAAAGTACCGGAA 272
Db      182  GTACAGAACGGCGTTAGACCTAGCGCTTGACAGAACCTCAGACCCAAAAGTACCGGAA 241
QY      273  GCATGTGTCGCGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 332
Db      242  GCATGTGTCGCGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 301
QY      333  ACTCTGGCAAGACAGTATGTTTACAGAAATCTTAAATAGTTTAAATAGTTTAAAGCCGC 392
Db      302  ACCTCTGGCAAGACAGTATGTTTACAGAAATCTTAAATAGTTTAAATAGTTTAAAGCCGC 361
QY      393  AGCAGCTGATTTTACACCCAGTTTACTAGAAAAGGAGGAGCAGTACTGACCTGAGTA 452
Db      362  AGCAGCTGATTTTACACCCAGTTTACTAGAAAAGGAGGAGCAGTACTGACCTGAGTA 421
QY      453  AAGGAAGGTGAAAAAGGAAGGACACTTCTACTATCTACAAAAAATCTCGAATGCAAT 512
Db      422  AAGGAAGGTGAAAAAGGAAGGACACTTCTACTATCTACAAAAAATCTCGAATGCAAT 481
QY      482  ATCAGAAAGATCTTATATAGTACAGGTGACAGATATGTCCTTAAAGGGGGTCTTAAAG 541
Db      513  ATCAGAAAGATCTTATATAGTACAGGTGACAGATATGTCCTTAAAGGGGGTCTTAAAG 572
QY      482  ATCAGAAAGATCTTATATAGTACAGGTGACAGATATGTCCTTAAAGGGGGTCTTAAAG 541
Db      573  AAAAGACATTTGCTTATAGTACCAACTGTGAGAGTGGCCAGTTTAAATATGACATCAAGCC 632
QY      542  AAAAGACATTTGCTTATAGTACCAACTGTGAGAGTGGCCAGTTTAAATATGACATCAAGCC 601
Db      633  CCATCTGGGGAGGAGACAGCAGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 692
QY      602  CCATCTGGGGAGGAGACAGCAGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 661
Db      693  CATTTGTCTACTACTGTTTGACAGAAAT 720
QY      662  CATTTGTCTACTACTGTTTGACAGAAAT 689
Db      662  CATTTGTCTACTACTGTTTGACAGAAAT 689

RESULT 8
B0780117 640 bp mRNA linear EST 26-JUL-2002
LOCUS      UI-R-FF0-cow-j-10-0-UI.s1 UI-R-FF0 Rattus norvegicus cDNA clone
DEFINITION UI-R-FF0-cow-j-10-0-UI 3', mRNA sequence.
ACCESSION B0780117
VERSION    B0780117.1 GI:21988589
KEYWORDS   EST.
SOURCE      Rattus norvegicus (Norway rat)
ORGANISM   Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sclurognathia; Muridae; Murinae;
            Rattus.
REFERENCE  1 (bases 1 to 640)
            Bernaldo,M.F., Lennon,G. and Soares,M.B.
            Normalization and subtraction: two approaches to facilitate gene
            discovery
JOURNAL    Genome Res. 6 (9), 791-806 (1996)
MEDLINE    97044477
PUBMED     8889548
COMMENT    Contact: Soares, MB
            Coordinated Laboratory for Computational Genomics
            University of Iowa
            375 Newton Road, 4156 MEBR, Iowa City, IA 52242, USA
            Tel: 319 335 8250
            Fax: 319 335 9565
            Email: bento-soares@uiowa.edu
            Tissue Procurement: Jeff Stevens
            cDNA library preparation: Dr. M. Bento Soares, University of Iowa

```

```

FEATURES
Source
location/Qualifiers
1..640
/organism="Rattus norvegicus"
/mol_type="mRNA"
/db_xref="taxon:10116"
/clone="UI-R-FF0-cow-j-10-0-UI"
/tissue_type="Mixed tissues"
/dev_stage="Adult"
/lab_host="DH10B (Life Technologies) (T1 phage resistant)"
/notes="Vector: pT73-Pac (Pharmacia) with a modified
polylinker; Site.1: EcoR I; Site.2: Not I; UI-R-FF0 is a
subtracted cDNA library containing the following tissue(s)
): Normal cartilage and SR-JWS tumor line. The
subtraction was made according to Bernaldo, Lennon and
Soares, Genome Research, 6:791-806, 1996. The
oligonucleotide used to prime the synthesis of
first-strand cDNA contains a library tag sequence that is
located between the Not I site and the (dri)18 tail. The
sequence tags for these libraries are: CTAAATGACG,
CATCTTGTA,
TAG-LIB-UI-R-FF0
TAG-TISSUE-rat SRC-JWS tumor line
TAG-SRO-CATCTTGTA"

BASE COUNT  208 a 121 c 152 g 159 t

Query Match      50.2%; Score 635.8; DB 13; Length 640;
Best Local Similarity 99.7%; Pred. No. 3.8e-74;
Matches 637; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY      33  TTTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 92
Db      2  TTTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 61
QY      93  TAACTGATAGCAAAAAAGATCAATGTAAAAACCTCAGCAATCTGCAACTGTCATTTGA 152
Db      62  TAACTGATAGCAAAAAAGATCAATGTAAAAACCTCAGCAATCTGCAACTGTCATTTGA 121
QY      153  AAAAAGTTTGTCTAGTGTGCGAAAAGGCCAACACTGTGTTCTTGCCAGTAGTTAGTT 212
Db      122  AAAAAGTTTGTCTAGTGTGCGAAAAGGCCAACACTGTGTTCTTGCCAGTAGTTAGTT 181
QY      213  GTACAGAACGGCGTTAGACCTAGCGCTTGACAGAACCTCAGACCCAAAAGTACCGGAA 272
Db      182  GTACAGAACGGCGTTAGACCTAGCGCTTGACAGAACCTCAGACCCAAAAGTACCGGAA 241
QY      273  GCATGTGTCGCGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 332
Db      242  GCATGTGTCGCGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 301
QY      333  ACTCTGGCAAGACAGTATGTTTACAGAAATCTTAAATAGTTTAAATAGTTTAAAGCCGC 392
Db      302  ACCTCTGGCAAGACAGTATGTTTACAGAAATCTTAAATAGTTTAAATAGTTTAAAGCCGC 361
QY      393  AGCAGCTGATTTTACACCCAGTTTACTAGAAAAGGAGGAGCAGTACTGACCTGAGTA 452
Db      362  AGCAGCTGATTTTACACCCAGTTTACTAGAAAAGGAGGAGCAGTACTGACCTGAGTA 421
QY      453  AAGGAAGGTGAAAAAGGAAGGACACTTCTACTATCTACAAAAAATCTCGAATGCAAT 512
Db      422  AAGGAAGGTGAAAAAGGAAGGACACTTCTACTATCTACAAAAAATCTCGAATGCAAT 481
QY      513  ATCAGAAAGATCTTATATAGTACAGGTGACAGATATGTCCTTAAAGGGGGTCTTAAAG 572
Db      482  ATCAGAAAGATCTTATATAGTACAGGTGACAGATATGTCCTTAAAGGGGGTCTTAAAG 541

```

```

OY 573 AAACACTGCTGTAAGTATGACACGTGTGAGATGGCCAGTAAATATGACTCAAGCC 632
DB 542 AAAACACTGCTGTAAGTATGACACGTGTGAGATGGCCAGTAAATATGACTCAAGCC 601
OY 633 CCATCTGGGAGGAGCAGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 671
DB 602 CCATCTGGGAGGAGCAGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 640

RESULT 9
CA339407/c 644 bp mRNA linear EST 04-NOV-2002
LOCUS NISL_1Y02409.y1 NCI_CGAP_Pf32 Rattus norvegicus cDNA clone
DEFINITION IMAGE:562185 5', mRNA sequence.
ACCESSION CA339407
VERSION CA339407.1 GI:24557505
KEYWORDS EST.
SOURCE Rattus norvegicus (Norway rat)
ORGANISM Rattus norvegicus
          Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
          Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
          Rattus.
REFERENCE 1 (bases 1 to 644)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
JOURNAL Tumor Gene Index
COMMENT Unpublished
          Contact: Robert Strausberg, Ph.D.
          Email: cgapbs-remail.nih.gov
          cDNA Library Preparation:
          cDNA Library Arrayed by: The I.M.A.G.E. Consortium/LLNL
          DNA Sequencing by: National Institutes of Health Intramural
          Sequencing Center (NISC)
          Clone distribution: NCI-CGAP clone distribution information can be
          found through the I.M.A.G.E. Consortium/LLNL at:
          InfoImage.llnl.gov
          Plate: LLM12447 row: G column: 18
          Seq primer: M13RP1 reverse primer (AB1).
          Location/Qualifiers
            1..644
            /organism="Rattus norvegicus"
            /mol_type="mRNA"
            /db_xref="taxon:10116"
            /clone="IMAGE:562185"
            /sex="male"
            /tissue_type="dorsal prostate"
            /dev_stage="adult, 14 month"
            /lab_host="DH10B (T1 phage-resistant)"
            /clone_host="NCI_CGAP_Pf32"
            /note="Organ: prostate; Vector: pCMV-SPORT6; Site_1: NCI;
            Site_2: Salt; Cloned unidirectionally. Primer: Oligo dT.
            Average insert size 2 kb. Constructed by Invitrogen. Note:
            this is a NCI_CGAP Library."
BASE COUNT 145 a 164 c 129 g 206 t
ORIGIN

Query Match 49.9%; Score 631.4; DB 14; Length 644;
Best Local Similarity 99.8%; Pred. No. 1.4e-73;
Matches 632; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

```

DB 464 GGTCTAGAGGGGGCGGCATCAATCATCATGATGTGTGTTACTCTGGCAAGACAGTATG 405
OY 353 TTTTCAGAAATATCTAAATATGTTTAAATGTTAAAGCCGACAGTGTCTTCTACACC 412
DB 404 TTTTCAGAAATATCTAAATATGTTTAAATGTTTAAAGCCGACAGTGTCTTCTACACC 345
OY 413 AGTTACTAGAAAACGAAAGGAAAGCACTACTACTAGTAAAGAAAGTGAAGAAAGGAA 472
DB 344 AGTTACTAGAAAACGAAAGGAAAGCACTACTACTAGTAAAGAAAGTGAAGAAAGGAA 285
OY 473 CGCACTTCTACTATCTACCAAAAATCTCCGAATGCATTTATCAGAAAAGATCTTATAGTA 532
DB 284 CCACACTTCTACTATCTACCAAAAATCTCCGAATGCATTTATCAGAAAAGATCTTATAGTA 225
OY 533 CAGGTCAGACATATTCCTGTTAAAGAGGGGCTCTTAAAGAAACACACTTGTATAGTAG 592
DB 224 CAGGTCAGACATATTCCTGTTAAAGAGGGGCTCTTAAAGAAACACACTTGTATAGTAG 165
OY 593 CAACCTGTGAGATGGCCACTTAAATATGACTCAACGCCCATCTGGGAGGAGCAGCA 652
DB 164 CAACCTGTGAGATGGCCACTTAAATATGACTCAACGCCCATCTGGGAGGAGCAGCA 105
OY 653 GGGGGAAGGGGGGCTCAAGAGACACTGATAGATCGGCATTTGTCATCTACTGTTTG 712
DB 104 GGGGGAAGGGGGGCTCAAGAGACACTGATAGATCGGCATTTGTCATCTACTGTTTG 45
OY 713 ACAGAAATTAACCGTTAAACACTTTACCCGCG 745
DB 44 ACAGAAATTAACCGTTAAACACTTTACCCGCG 12

RESULT 10
BI283790 638 bp mRNA linear EST 19-JUL-2001
LOCUS UI-R-CW05-cch-g-03-0-UI.s1 UI-R-CW05 Rattus norvegicus cDNA clone
DEFINITION UI-R-CW05-cch-g-03-0-UI 3', mRNA sequence.
ACCESSION BI283790
VERSION BI283790.1 GI:14935880
KEYWORDS EST.
SOURCE Rattus norvegicus (Norway rat)
ORGANISM Rattus norvegicus
          Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
          Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
          Rattus.
REFERENCE 1 (bases 1 to 638)
AUTHORS Bonaldo,M.F., Lennon,G. and Soares,M.B.
TITLE Normalization and subtraction: two approaches to facilitate gene
JOURNAL discovery
MEDLINE Genome Res. 6 (9), 791-806 (1996)
PUBMED 97044477
COMMENT 8889548
          Contact: Soares, MB
          Coordinated Laboratory for Computational Genomics
          University of Iowa
          375 Newton Road, 4156 MEBRF, Iowa City, IA 52242, USA
          Tel: 319 335 8250
          Fax: 319 335 9565
          Email: bento-soares@uiowa.edu
          The sequence contained an oligo-dT track that was present in the
          oligonucleotide that was used to prime the synthesis of first
          strand cDNA and therefore this may represent a bonafide poly A
          tail. The sequence tag present in the cDNA between the NotI site
          and the oligo-dT track served to identify it as a clone from the
          non-normalized rat aorta pool library cDNA library preparation:
          M.B. Soares Lab Clone distribution: clones will be available
          through Research Genetics (www.resgen.com)
          Seq primer: M13 Forward
          POLA=Yes
          Location/Qualifiers
            1..638
            /organism="Rattus norvegicus"
            /mol_type="mRNA"
FEATURES
source

```





```

OY 153 AAAAGTTGTTCTAGTGTGAGAAAGCCCAACACTGTGTCTTCCAGTGAATAGGTT 212
    |||||||
DB 121 AAAAGTTGTTCTAGTGTGAGAAAGCCCAACACTGTGTCTTCCAGTGAATAGGTT 180
OY 213 GTACGAAGGCGGTAGACACTAGGCGTTCAGCAACCTCAGACCCAAAGGTACCAGAA 272
    |||||||
DB 181 GTACGAAGGCGGTAGACACTAGGCGTTCAGCAACCTCAGACCCAAAGGTACCAGAA 240
OY 273 GCATGTGTCCGCGGTGAGGTGAGGTGAGAGGGGCGGCATCATCATGACAGTGTGGT 332
    |||||||
DB 241 GCATGTGTCCGCGGTGAGGTGAGGTGAGAGGGGCGGCATCATCATGACAGTGTGGT 300
OY 333 ACTGTGCAAGACACTGATGTTTTCAGAAATCTTAAATATGTTTAAAACTGTAAAGCCGC 392
    |||||||
DB 301 ACTGTGCAAGACACTGATGTTTTCAGAAATCTTAAATATGTTTAAAACTGTAAAGCCGC 360
OY 393 AGCAGCTGATTTCTACCCAGTCTACTAGAAAACGAAGGAAGCACTAGTCACTGAGTA 452
    |||||||
DB 361 AGCAGCTGATTTCTACCCAGTCTACTAGAAAACGAAGGAAGCACTAGTCACTGAGTA 420
OY 453 AAGGAAGGTGAAAACAGAGACGCACTTCTACTATCTACCAAAAAATCTCCGATGCAAT 512
    |||||||
DB 421 AAGGAAGGTGAAAACAGAGACGCACTTCTACTATCTACCAAAAAATCTCCGATGCAAT 480
OY 513 ATCAGAAAGATCTTATAGTACAGTCAAGTATATTGCTGTTAAGAGGGGTCCTAAAG 572
    |||||||
DB 481 ATCAGAAAGATCTTATAGTACAGTCAAGTATATTGCTGTTAAGAGGGGTCCTAAAG 540
OY 573 AAAAGCACTTGTAGTATAGCAACTGTGAGATGGCCAGTTTAAATATGACTCAACGCC 632
    |||||||
DB 541 AAAAGCACTTGTAGTATAGCAACTGTGAGATGGCCAGTTTAAATATGACTCAACGCC 600
OY 633 CCATCTGGGAGGAGCAGCAGGCGGGAAGGGCG 664
    |||||||
DB 601 CCATCTGGGAGGAGCAGCAGGCGGGAAGGGCG 632
```

RESULT 12  
B0780658 636 bp mRNA linear EST 26-JUL-2002  
LOCUS UI-R-FP0-cpc-c-22-0-UI.s1 UI-R-FP0 Rattus norvegicus cDNA clone  
DEFINITION UI-R-FP0-cpc-c-22-0-UI 3', mRNA sequence.

ACCESSION B0780658  
VERSION B0780658.1 GI:21989130  
KEYWORDS EST.  
SOURCE Rattus norvegicus (Norway rat)  
ORGANISM Rattus norvegicus

REFERENCE 1 (bases 1 to 636)  
AUTHORS Bonaldo,M.F., Lennon,G. and Soares,M.B.  
TITLE Normalization and subtraction: two approaches to facilitate gene  
discovery

JOURNAL Genome Res. 6 (9), 791-806 (1996)  
MEDLINE 97044477  
PUBMED 8889548  
COMMENT Contact: Soares, MB  
Coordinated Laboratory for Computational Genomics  
University of Iowa  
375 Newton Road, 4156 MEBRF, Iowa City, IA 52242, USA  
Tel: 319 335 8250  
Fax: 319 335 9565  
Email: bento-soares@uiowa.edu  
Tissue Procurement: Jeff Stevens  
cDNA Library preparation: Dr. M. Bento Soares, University of Iowa  
cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa  
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa  
Clone Distribution: DISTRIBUTION: Researchers may obtain clones  
from Research Genetics (www.resgen.com).  
Seq primer: M13 FORWARD  
POLA=yes.

FEATURES  
Location/Qualifiers

```

source 1. 636
        /organism="Rattus norvegicus"
        /mol_type="mRNA"
        /db_xref="taxon:10116"
        /clone="UI-R-FP0-cpc-c-22-0-UI"
        /tissue_type="Mixed tissues"
        /dev_stage="Adult"
        /lab_host="DH10B (Life Technologies) (T1 phage resistant)"
        /note="Vector: pUT73-Pac (Pharmacia) with a modified
        polylinker; Site_1: EcoR I; Site_2: Not I; UI-R-FP0 is a
        subtracted cDNA library containing the following tissue(s)
        : Normal cartilage and SR-TWS Tumor Line . The
        subtraction was made according to Bonaldo, Lennon and
        Soares, Genome Research, 6:791-806, 1996. The
        oligonucleotide used to prime the synthesis of
        first-strand cDNA contains a library tag sequence that is
        located between the Not I site and the (dt)18 tail. The
        sequence tags for these libraries are: CTATGTGACG,
        CATCTGTGA.
        TAG_LIB=UI-R-FP0
        TAG_TISSUE=cartilage
        TAG_SEQ=CTATGTGACG"

BASE COUNT 210 a 120 c 148 g 158 t

Query Match 49.3%; Score 623.8; DB 13; Length 636;
Best Local Similarity 98.9%; Pred. No. 1.4e-72;
Matches 628; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

OY 33 TTTTGTGTTTATGATTCAGATTTTATTAAGTATCATATGCAAAACATATGTC 92
    |||||||
DB 2 TTTTGTGTTTATGATTCAGATTTTATTAAGTATCATATGCAAAACATATGTC 61
OY 93 TTAAGTATTCAGATTTTATTAAGTATTCAGATTTTATTAAGTATTCAGATTTTATTAAG 152
    |||||||
DB 62 TTAAGTATTCAGATTTTATTAAGTATTCAGATTTTATTAAGTATTCAGATTTTATTAAG 121
OY 153 AAAAGTTGTTCTAGTGTGAGAAAGCCCAACACTGTGTCTTCCAGTGAATAGGTT 212
    |||||||
DB 121 AAAAGTTGTTCTAGTGTGAGAAAGCCCAACACTGTGTCTTCCAGTGAATAGGTT 180
OY 213 GTACGAAGGCGGTAGACACTAGGCGTTCAGCAACCTCAGACCCAAAGGTACCAGAA 272
    |||||||
DB 181 GTACGAAGGCGGTAGACACTAGGCGTTCAGCAACCTCAGACCCAAAGGTACCAGAA 240
OY 273 GCATGTGTCCGCGGTGAGGTGAGGTGAGAGGGGCGGCATCATCATGACAGTGTGGT 332
    |||||||
DB 241 GCATGTGTCCGCGGTGAGGTGAGGTGAGAGGGGCGGCATCATCATGACAGTGTGGT 300
OY 333 ACTGTGCAAGACACTGATGTTTTCAGAAATCTTAAATATGTTTAAAACTGTAAAGCCGC 392
    |||||||
DB 301 ACTGTGCAAGACACTGATGTTTTCAGAAATCTTAAATATGTTTAAAACTGTAAAGCCGC 360
OY 393 AGCAGCTGATTTCTACCCAGTCTACTAGAAAACGAAGGAAGCACTAGTCACTGAGTA 452
    |||||||
DB 361 AGCAGCTGATTTCTACCCAGTCTACTAGAAAACGAAGGAAGCACTAGTCACTGAGTA 420
OY 453 AAGGAAGGTGAAAACAGAGACGCACTTCTACTATCTACCAAAAAATCTCCGATGCAAT 512
    |||||||
DB 421 AAGGAAGGTGAAAACAGAGACGCACTTCTACTATCTACCAAAAAATCTCCGATGCAAT 480
OY 513 ATCAGAAAGATCTTATAGTACAGTCAAGTATATTGCTGTTAAGAGGGGTCCTAAAG 572
    |||||||
DB 481 ATCAGAAAGATCTTATAGTACAGTCAAGTATATTGCTGTTAAGAGGGGTCCTAAAG 540
OY 573 AAAAGCACTTGTAGTATAGCAACTGTGAGATGGCCAGTTTAAATATGACTCAACGCC 632
    |||||||
DB 541 AAAAGCACTTGTAGTATAGCAACTGTGAGATGGCCAGTTTAAATATGACTCAACGCC 601
OY 633 CCATCTGGGAGGAGCAGCAGGCGGGAAGGGCGCT 667
    |||||||
DB 602 CCATCTGGGAGGAGCAGCAGGCGGGAAGGGCGCT 636
```

RESULT 13	615 bp	mrna	linear	EST 12-MAR-2001
LOCUS	BC380448			
DEFINITION	UI-R-CT0-b1x-d-05-0-UI.s1 UI-R-CT0 Rattus norvegicus cDNA clone			
ACCESSION	BC380448	UI-R-CT0-b1x-d-05-0-UI 3', mRNA sequence.		
VERSION	BC380448.1	GI:1304920		
KEYWORDS	EST.			
SOURCE	Rattus norvegicus (Norway rat)			
ORGANISM	Rattus norvegicus			

REFERENCE AUTHORS TITLE	JOURNAL MEDLINE PUBMED	COMMENT
1 (bases 1 to 615) Ronalds,M.F., Lennon,G. and Soares,M.B. Normalization and subtraction: two approaches to facilitate gene discovery	Genome Res. 6 (9), 791-806 (1996) 97044477 8889548	
Contact: Soares, MB		

Contact: Soares, MB  
Coordinated Laboratory for Computational Genomics  
University of Iowa  
375 Newton Road, 4156 MEBRF, Iowa City, IA 52242, USA  
Tel: 319 335 8250  
Fax: 319 335 9565  
Email: bento-soares@uiowa.edu

The sequence contained an oligo-dT track that was present in the oligonucleotide that was used to prime the synthesis of first strand cDNA and therefore this may represent a bonafide poly A tail. The sequence tag present in the cDNA between the NotI site and the oligo-dT track served to verify it as a clone from the normalized rat brain pool library cDNA library Preparation: M.B. Soares Lab Clone distribution: clones will be available through Research Genetics ([www.resgen.com](http://www.resgen.com))

Seq primer: M13 Forward  
SOLTA-tes.

FEATURES	Location/Qualifiers
source	1. .615

```

SOURCE
1. 013
/organism="Rattus norvegicus"
/mol_type="mRNA"
/strain="Sprague-Dawley"
/db_xref="taxon:10116"
/clone="UT-R-CT0-btx-d-05-0-UT"
/dev_stage="ADULT"
/lab_host="DH10B (Life Technologies)"
/clone_lib="UT-R-CT0"
/notes="Vector: pRTT3D-Pac (Pharmacia) with a modified
polylinker. Site_1: Not I; Site_2: Eco RI; The UT-R-CT0
library is a normalized library constructed from the
following rat brain tissues: embryonic day 17, embryonic
day 19, embryonic day 21, adult day 1, adult day 12, adult
day 75, adult day 200. For a detailed description of the
library from which this clone was derived, please visit
our web site at ratest.eng.uiowa.edu. The subtraction has
been previously described in (Bonaldo, Lennon and Soares,
1996)
TAG_lib=UT-R-CT0
TAG_TISSUE=rat brain pool
TAG_SEQ=ACTCTC"

```

Qy	33	TTTTTTTTTTTTTTTATGATTCAGAGATTATTAACTCATACATCGAAACACTATCGC	92
Db	2	TTTTTTTTTTTTTTTTTANGATTCAGAGATTATTAACTCATACATCGCAAAACACTATCGC	61
		Matches 611; Conservative 0; Mismatches 3; Indels 0; Gaps 0;	
		48.2%; Score 609.8; DB 10; Length 615;	
		Best Local Similarity 99.5%; Pred. NO. 9.8e-71;	

OY	93	TAATGCATTGCAAAAGATCAATATGTAAAACTCCACATTCGTGCAAGCTGCATATTTGA	152
Db	62	TAATGCATTGCAAAAGATCAATATGTAAAACTCCACATTCGTGCAAGCTGCATATTTGA	121
OY	153	AAAAAGTTGTTCTAGTGTCGAAAGGCCAACACTGTGTCTTGCCAGAGTCTAGATT	212
Db	122	AAAAAGTTGTTCTAGTGTCGAAAGGCCAACACTGTGTCTTGCCAGAGTCTAGATT	181
OY	213	GTACAGAACGGCGTTAGCACTAGCGCTTGACAGAACTCACAGACCCAAAGTACCGGAA	272
Db	182	GTACAGAACGGCGTTAGCACTAGCGCTTGACAGAACTCACAGACCCAAAGTACCGGAA	241
OY	273	GCATGTGCCCGGGGGGAGGCTGAGAGGGGGGGCATCAATCAATGACAGTGTGCT	332
Db	242	GCATGTGCCCGGGGGGAGGCTGAGAGGGGGGGCATCAATCAATGACAGTGTGCT	301
OY	333	ACTGTGGCAAGACAGTAGTGTCTCAGAAATCTTAAATAGTTTAAAACTGTAAAGCCGC	392
Db	302	ACTGTGGCAAGACAGTAGTGTCTCAGAAATCTTAAATAGTTTAAAACTGTAAAGCCGC	361
OY	393	AGCAGCTGATTTCTACACCCAGTTACTAGAAAAGAAAGGAAGCACTAGTCAGTGAATA	452
Db	362	AGCAGCTGATTTCTACACCCAGTTACTAGAAAAGAAAGGAAGCACTAGTCAGTGAATA	421
OY	453	AAGGAAGGTGAAAACAGSAGACGCACTCTTACTATCTACCAAAAAAATCTCCGATGCATT	512
Db	422	AAGGAAGGTGAAAACAGSAGACGCACTCTTACTATCTACCAAAAAAATCTCCGATGCATT	481
OY	513	ATCAGAAAAGATCTTATAGTACAGGTGCAGACATATTGCTCGTTAAGAAAGGGGCTCTTAAG	572
Db	482	ATCAGAAAAGATCTTATAGTACAGGTGCAGACATATTGCTCGTTAAGAAAGGGGCTCTTAAG	541
OY	573	AAAAAGCACTTGCTAAGTATGCAACTGTGAGAGTGGCCAGTTTAAATATGCACTACAGCC	632
Db	542	AAAAAGCACTTGCTAAGTATGCAACTGTGAGAGTGGCCAGTTTAAATATGCACTACAGCC	601
OY	633	CCATCTGGGGAGGC	646
Db	602	CCATCTGGGGAGGC	615

```

RESULT 14
LOCUS      B0514423/c
DEFINITION AGENCOURT 10119669 NIH MGC.134 Mus musculus cDNA clone
ACCESSION  B0514423
VERSION     B0514423.1
KEYWORDS    GI:22821949
SOURCE      EST.
            Mus musculus (house mouse)
ORGANISM    Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Scurionathi; Muridae; Murinae; Mus.
REFERENCE   1 (bases 1 to 978)
            NIH-MGC http://mgc.nci.nih.gov/.
            National Institutes of Health, Mammalian Gene Collection (MGC)
            Unpublished
AUTHORS     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs@email.nih.gov
TITLES      Tissue Procurement: Dr. David Rowe
            cDNA Library Preparation: Invitrogen Corp
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LMNL)
            DNA Sequencing by: Agencourt Bioscience Corporation
            Clone distribution: MGC clone distribution Information can be
            found through the I.M.A.G.E. Consortium/LMNL at:
            http://image.lnl.gov
            Plate: L14M14080 row: 0 column: 14
            High quality sequence start: 17
            High quality sequence stop: 714.
FEATURES    location/qualifiers
            1..978
            /organism="Mus musculus"
            /mol_type="mRNA"

```

			/db_xref="taxon:10090"	
			/clone_image="6511045"	
			/russ_type="undifferentiated limb"	
			/lab_host="DH10B (phage-resistant)"	
			/clone_lib="NH_MGC_134"	
			/note="Vector: pCMV-SPORT6.1.ccdB; Site.1: EcoRV; Site.2: NotI; Cloned unidirectionally. Primer: Oligo dT. Average insert size 1.7 kb. Constructed by Resgen, Invitrogen Corp. Note: This is a NH_MGC library."	
BASE COUNT	249 a	224 c	201 g	301 t
ORIGIN				3 others
Query Match	47.3%	Score 599.2;	DB 13;	Length 978;
Matches Local Similarity	85.9%	Pred. No.1.9e-69;	Mismatches 94;	Indels 32; Gaps 8;
Bases	767;	Conservative 0;	Partial 96;	
QY	402	TTCCTACACCAGTACTAGAAAAGGAGAGCAGTAGTGAG--CTGAGTAAGAAGAA	459	
Db	953	TTGGACCAACCTAAGTTGTATTAGAACCTAAGGAAGCCACTAAGTCTCGATGAAGTAGG	894	
QY	460	GTCGAAACAGGAGCAGCTTCTACT--ATCTACCAAAAAATCTCCGAA-----TGCAT	511	
Db	893	GCGAAAACAGAGAGCGCACTTCTACTGTAGCTACCAAAAAAACAATAAAATCTWTGGCT	834	
QY	512	TATCAGAAAGATCTTATAGTAGCAGGTGACGACATATGTGCTGTAAAGAGGGGCTCTAAA	571	
Db	833	TATCAGAAAGATCTTATAGTAGCAGGTGACGACATATGTGCTGTAAAGAGGGG--TCTAAA	776	
QY	572	GAAGAAGACTGTCTAAGTTAGCACTGAGAGATGAGCCAGTTTAAATATGACCTCACGC	631	
Db	775	GAAGAAGACTGTCTAAGTTAGCACTGAGAGATGAGCCAGTTTAAATATGACCTCACGC	716	
QY	632	CCCATCTGGGGAGGAGCAGCAGGGGGGGAAGGGGGCTCAAG-----AGAGACAC	679	
Db	715	CCCATCTGGGGAGGAGCAGCAGGTGAGGGGTGGGGGGGCCGAAGGGATGCTCAAGAGACAC	656	
QY	680	TGATTAAGATCGGCCATTTTGTATCTACTGTTTGACAGAAATTAACCGTTAAAAAGCTTTA	739	
Db	655	TGATTAAGATCGGCCATTTTGTATCTACTGTTTGACAGAAATTAACCGTTAAAAAGCTTTA	596	
QY	740	CCCGTGACACTTATTCAGTTGAATTAATCTCCAGTAGTAAGTAGTGTAATTAATCTCT	799	
Db	595	CCCGTGACACTTATTCAGTTGAATTAATCTCCAGTAGTAAGTAGTGTAATTAATCTCT	536	
QY	800	ACTTCATATTAAGTCAAAATTAAGTCTGTCTCTCTCTTGATGAGGTGCTGTCTTTCACACATCC	859	
Db	535	ACTTCATATTAAGTCAAAATTAAGTCTGTCTCTCTCTTGATGAGGTGCTGTCTTTCACACATCC	478	
QY	860	ACCCAGCAGCACCCAGCAGCTAGGAACAGAAATCTTGCTTTAGAGGCACACAGAGGCCAGAG	919	
Db	477	ACCCAGCAGCACCCAGCAGCTAGGAACAGAAATCTTGCTTTAGAGGCACACAGAGGCCAGAG	418	
QY	920	TTCTGTTCAAAGCCCTGCGAAGCGCGGTGAGTGTGTAAGTAAGTCACTATGAATTC	979	
Db	417	TTCTGTTCAAAGCTGCGAAGAGCTAGTGTGCTGTGCTTTTGAAGAAATCACTATGAATTC	358	
QY	980	AAGAGCAGAGCTTTACACCCCATC--GTGACGTACGTACAAAGTTACGTAAATGAGCA	1036	
Db	357	AAGAGCAGAGCTTTACACACATCAACACGTGACGTAAGTACAAAGTTACGTAAATGAGCA	298	
QY	1037	TGGGCTGATTAAGTTACAGGTGCTGTACATGCGACGCTGTCAATTAAGAGGCTGTGCTGTG	1096	
Db	297	TGGGCTGATTAAGTTACAGGTGCTGTACATGCGAACGTGTCTTAAGAGGCTGTGCTGTG	238	
QY	1097	TCACACGGTCTGGGAGGTAGGGGAGGTCTCACCCCCGAGGCCCAAGAGCTGCAAGTCTTC	1155	
Db	237	TCACACGGTCTGGGAACTCTGGAAGGGCTCTCACCCCCGAGGCCCAAGAGCTGCGGTCTTC	178	
QY	1157	TTAAGAGCAAA--GTCTCTCACAGCTTTAGTGTATGCTTTCACACAGCAAGCACT	1213	
Db	177	TTACACAGAGAGTCTCTCAGAGAGGTTATGCTTATGCTTTCACACAGCAAGCAACT	118	
QY	1214	TAGTTCCACAAAGCTATTTTGGCAATTCCTTAATCTGACAGAAATAGGGGATTTT	1266	

Db	117	TAGTTCAGAGGATTTGGCAATTTTAAATCTGACGAGAAGATGGGAGATTT	65
RESULT 15	CA318577	826 bp	mus musculus
LOCUS	CA318577/c		
DEFINITION	UI-M-FW0-cdr-O-14-0-UI_r1 NIH_BMP_FW0	mus musculus	CDNA clone
ACCESSION	CA318577		
VERSION	CA318577.1	GI:24536701	
KEYWORDS	EST.		
ORGANISM	Mus musculus		
REFERENCE	1	(bases 1 to 826)	
AUTHORS	NIH-MGC	http://mgc.nci.nih.gov/	
TITLE	National Institutes of Health		
JOURNAL	Unpublished		
COMMENT	Contact: Robert Strausberg, Ph.D. Email: cga@bbs-remail.nih.gov Tissue Procurement: Dr. Jim Lin, University of Iowa CDNA Library preparation: Dr. M. Bento Soares, University of Iowa CDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa DNA Sequencing by: Dr. M. Bento Soares, University of Iowa Clone Distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/ILNML at: <a href="http://image.jiml.gov">http://image.jiml.gov</a> This clone was contributed by the Brain Molecular Anatomy Project (BMAP)		
FEATURES	Seq primer: pYX-5.		
source	Location/Qualifiers		
	1..826		
	/organism="Mus musculus"		
	/mol_type="mRNA"		
	/strain="C57BL/6"		
	/db_xref="taxon:10090"		
	/clone="IMAGE: 6813639"		
	/tissue="type="whole brain"		
	/dev_stage="embryo 13.5, 14.5, 16.5, 17.5, 19.5"		
	/lab_host="DH10B (T1 phage resistant)"		
	/clone_11b="NIH_BMAP_FW0"		
	/note="Organ: Brain; Vector: pYX-Asc; Site: 1: EcoR I; Site: 2: Not I; The library was constructed according to Bonaldi, Lennon and Soares, Genome Research, 6:791-806, 1996. Denatured RNA was size fractionated on a 1% agarose gel. First strand cDNA synthesis was primed with oligo-dT primer containing a Not I site. Double strand cDNA was size selected according to mRNA size fraction, ligated with EcoR I adaptor, digested with NotI and then cloned directionally into pYX-Asc vector. The library tag is ACCGAGACAG. This library was created for the University of Iowa Brain Anatomy Project (BMAP). 'Gene discovery in the developing mouse nervous system', supported by National Institute of Mental Health (NIMH), Hemlin Chlin, Ph.D., program coordinator."		
BASE COUNT	226 a	195 c	162 g
ORIGIN	242 t	1 others	
Query Match	44.8%	Score 567	DB 14
Best Local Similarity	89.7%	Pred. No. 3.3e-65	
Matches 602	Conservative 0	Mismatches 56	Indels 22
	Gaps 6		
QY	525	TATAGTACAGGTCAGACATATTCCTGCTTAAGAAGGGGCTCTAAAGAAACACTTGC	584
DB	826	TATAGTACAGGTCAGACATATTCCTGCTTAAGAAGGGGCTCTAAAGAAACACTTGC	768
OY	585	TATAGTACAGGTCAGACATATTCCTGCTTAAGAAGGGGCTCTAAAGAAACACTTGC	644
Db	767	TATAGTACAGGTCAGACATATTCCTGCTTAAGAAGGGGCTCTAAAGAAACACTTGC	709

```
QY 645 GGACAGCAGGGGGAGGGGGCTCAAG-----AGAGCACACTGATAGATCGGC 692
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 708 GGACGGCAGTGGAGGGGGGGCGGAGAGATGCTCAAGAGACACTGATTAAGATCGGC 649
QY 693 CATTTGTCATCTACTGTTTGACAGAAATTAACGTTAAAAAGCTTACCCGTGACACTTT 752
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 648 CATTTGTCACCTACTGTTTGACAGAAATTAACGTTAAAAAGCTTACCCGTGACACTTT 589
QY 753 TATTAGTTGAATTAATCTCCATGTAACAATGTAAGTAAATTAATCTACTTCAATTAAT 812
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 588 TATTCACTTGAATTAATCAATGTAACAATGTAAGTAAATTAATCTACTTCAATTAAT 529
QY 813 CAAATAATCTGTCGTCTCCCTTTATGACGTCGTGTTCAACACACTCCACCAGCACACC 872
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 528 CAAATAATCTGTCGTCTCCCTTTATGACGTCGTGTTCAACACACTCCACCAGCACACC 471
QY 873 ACGACTAGAGAAAGAAATTAATCTCGTTAGAGGCAACACAGAGCCAGAGTTCTGTTCAAAGC 932
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 470 ACAACTAGAGAAAGAAATTAATCTCATTTAGAGGCAACACAGAGCCAGAGTTCTGTTCAAAGT 411
QY 933 CTGCAGAGCCGGTCAAGCTGGTATTTTAGAGAACTCACTATGAATCAAGAGAGCAGAGCT 992
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 410 CTGCAGAGCCGGTCAAGCTGGTATTTTAGAGAACTCACTATGAATCAAGAGAGCAGAGCT 351
QY 993 GTTACACCCCATC---GTGAGGTACAGTACAAAGTTAGTAAATGAGCATGGGCTGATTAAGT 1049
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 350 GTTACACCTCATCACTGTGACGTACAGTACAAAGTTAGTAAATGAGCATGGGCTGATTAAGT 291
QY 1050 TACAGGTGCGTTACATGTCAGCGTGTCAATTAAGAGAGGCTGTGCTGTCTCACACGTCCTGG 1109
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 290 TACAGGTGCGTTACATGTCAGCGTGTCAATTAAGAGAGGCTGTGCTGTCTCACACGTCCTGG 231
QY 1110 GAGCTACGGAGGGGTCTGCACCCCTGAGCCGAGAGCTGCAAGTCTTCTTAAGGACAAA-- 1167
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 230 GAACTCTGGAAGGGTCTGCACCCCTGAGCCGAGAGCTGCGGCTTCTCTTAGCAACAGAGAG 171
QY 1168 -GTCCTCAACAGCTTAGTGTGTTACGTTCTCTGACACAAGCAACTTACTTCAAGAGT 1226
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 170 TCTCTCAAGAGAGTTTGTGCTTAGTGTCTCTGACACAAGCAACTTACTTCAAGAGT 111
QY 1227 ATTTTGGCAATTTCTTAATCTGAGCAAGAAATAGGGGATTTT 1266
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 110 ATTTTGGCAATTTTCTTAATCTGAGCAAGAAATAGGGGATTTT 71
```

Search completed: August 25, 2003, 23:35:31  
Job time : 2641.53 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 25, 2003, 21:58:43 ; Search time 92.8689 Seconds  
(without alignments)  
6016.990 Million cell updates/sec

Title: US-09-717-321A-2  
Perfect score: 1266  
Sequence: 1 ttttttttttttttttcaaa.....gagcaagatagggatttt 1266

Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 569978 seqs, 220691566 residues

Total number of hits satisfying chosen parameters: 1139956

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued\_Patents\_NA.\*  
1: /cgn2\_6/ptodata/1/ina/5A.COMB.seq:\*  
2: /cgn2\_6/ptodata/1/ina/5B.COMB.seq:\*  
3: /cgn2\_6/ptodata/1/ina/6A.COMB.seq:\*  
4: /cgn2\_6/ptodata/1/ina/6B.COMB.seq:\*  
5: /cgn2\_6/ptodata/1/ina/PT05.COMB.seq:\*  
6: /cgn2\_6/ptodata/1/ina/Backfiles1.seq:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	187.2	14.8	605	3	US-09-385-982-483
2	68	5.4	588	3	US-09-385-982-133
3	54.2	4.3	7218	1	US-08-232-463-14
4	49.8	3.9	5852	1	US-07-867-106-2
5	47.8	3.8	2218	4	US-09-205-258-103
6	46.8	3.7	760	4	US-09-205-258-232
7	45.6	3.6	38844	4	US-09-734-675-3
8	45	3.6	1882	3	US-09-370-253-1
9	44.6	3.5	991	3	US-08-924-747-25
10	44.6	3.5	991	3	US-09-247-373B-25
11	43.2	3.4	615	4	US-09-126-715-25
12	43.2	3.4	615	4	US-09-105-542A-2
13	43	3.4	1098	3	US-09-248-335-35
14	42.2	3.3	1507	4	US-09-453-323-1
15	42.2	3.3	8100	4	US-09-554-337-4
16	42.2	3.3	11517	1	US-07-920-281C-1
17	42.2	3.3	11517	1	US-08-466-277-1
18	42.2	3.3	15538	4	US-09-554-337-1
19	41.8	3.3	1332	2	US-09-057-762-1
20	41.8	3.3	1332	3	US-08-326-119A-1
21	41.6	3.3	358	2	US-08-721-488-9
22	41.6	3.3	1342	4	US-09-489-847-89
23	41.6	3.3	1929	3	US-09-146-950-1
24	41.4	3.3	501	4	US-09-601-198-170
25	41.2	3.3	1641	1	US-08-300-903A-8
26	41.2	3.3	1641	1	US-08-988-197-8
27	41.2	3.3	1898	1	US-08-342-411A-1

C 28	41	3.2	966	1	US-08-514-014-7	Sequence 7, Appl
C 29	41	3.2	966	2	US-08-833-823-7	Sequence 7, Appl
C 30	40.8	3.2	1117	3	US-09-247-373B-33	Sequence 33, Appl
C 31	40.8	3.2	1878	4	US-09-465-558-39	Sequence 39, Appl
C 32	40.4	3.2	1736	3	US-09-182-816-22	Sequence 22, Appl
C 33	40.4	3.2	1736	3	US-09-182-816-22	Sequence 22, Appl
C 34	40.4	3.2	1736	3	US-09-471-528-22	Sequence 22, Appl
C 35	40.4	3.2	1736	3	US-09-471-528-22	Sequence 22, Appl
C 36	40.4	3.2	1736	3	US-09-634-530-22	Sequence 22, Appl
C 37	40.4	3.2	1736	3	US-09-634-530-22	Sequence 22, Appl
C 38	40.2	3.2	289	1	US-08-341-568-3	Sequence 24, Appl
C 39	40.2	3.2	289	2	US-08-911-020-3	Sequence 3, Appl
C 40	40.2	3.2	859	1	US-08-345-756-8	Sequence 8, Appl
C 41	40.2	3.2	859	1	US-08-345-756-8	Sequence 8, Appl
C 42	40.2	3.2	2039	1	US-08-625-198-5	Sequence 5, Appl
C 43	40.2	3.2	2039	1	US-08-625-198-5	Sequence 5, Appl
C 44	40	3.2	2246	4	US-09-363-708-3	Sequence 3, Appl
C 45	40	3.2	2246	4	US-09-083-587-3	Sequence 3, Appl

## ALIGNMENTS

RESULT 1	US-09-385-982-483
Sequence 483, Application US/09385982	
Patent No. 6262334	
GENERAL INFORMATION:	
APPLICANT: ENDEGE, WILSON O., ET AL.	
TITLE OF INVENTION: NOVEL HUMAN GENES AND GENE EXPRESSION	
FILE REFERENCE: CCDDNA-260XX	
CURRENT APPLICATION NUMBER: US/09/385, 982	
EARLIER FILING DATE: 1999-08-30	
EARLIER APPLICATION NUMBER: 09/328,111	
EARLIER FILING DATE: 1999-06-08	
EARLIER APPLICATION NUMBER: 60/117,393	
EARLIER FILING DATE: 1999-01-27	
EARLIER APPLICATION NUMBER: 60/098,639	
EARLIER FILING DATE: 1998-08-31	
NUMBER OF SEQ ID NOS: 544	
SOFTWARE: FASTSEQ for Windows Version 3.0	
SEQ ID NO 483	
LENGTH: 605	
TYPE: DNA	
ORGANISM: Homo sapiens	
FEATURE:	
NAME/KEY: misc_feature	
LOCATION: (1)...(605)	
OTHER INFORMATION: n = A,T,C or G	
US-09-385-982-483	
Query Match	14.8%; Score 187.2; DB 3; Length 605;
Best Local Similarity	65.5%; Pred. No. 3.3e-41;
Matches 348; Conservative	0; Mismatches 158; Indels 25; Gaps 5;
OY	265 TACCGGAGGAGGTCGCGCGGCGGAGGTCGTAGAGGGGGGATCAATACATGCA 324
DB	74 TACAGGAGCGGTCTCCGCGGCGGAGGTAAGA-GGGTCAGTATGCTCAAGTCA 132
OY	325 GTGTGGTACTCTGCGAGACAGTATGTTTCAGAA-----TATCTCA 367
DB	133 GTGTGGTACTCTGCGAGACAGTATGTTTCAGAA-----TATCTCA 192
OY	368 AATGCTTTAAACCTGTAAGCGCAGCAGTATTTACACCCAGTTACTAGAAAAG 427
DB	193 AATATTTTAAACCTGTAAGCGCAGCAGTATTTACACCTAGTTACTAGAAAAG 252
OY	428 AAGGAGGAGCAGTACAGTGTGAGTGAAGGAGGAGGAGGAGGAGGAGGAGGAG 487
DB	253 AAGGAGGAGCAGTATTTAGCTGTGAATA-----AAGTACAGTGAAGGAGGAGGAG 307
OY	488 TACCAAAAAAATCTCGAATGATTTATGAGAAAGATCTTATAGTACAGGTACAGATAT 547

Db	308	CGACAAAAAAACCTTCTTAATGCATTATCAGAAAGATTTTATTAATACAAAG -GAGGCATATT	366
Oy	548	GCATGTTAAGAAGGGGGCTCTAAGAAAAAGCACTGCTAATTTGCAACTGTGAGATGG	607
Db	367	GCATCATCAGAAGGGTTCTTATTAAGAAAAAGCACTTACTAATTGGCCACTTAACGAAACAA	428
Oy	608	CCAGTTTAAATATGCACTAACGCCCATCTGCGGAGGGACAGACAGGGGGCAAGGGGGCT	667
Db	427	CCAGTTTAAAGATTAATTAATGCCCCCATTTGGGGANCAATGGAGGGTTTAAACANNAAN	486
Oy	668	CAAGAGAGACACTGATTAAGATGGGCATTTGTC -ATTACTGTTTGGACAGAAATTAACG	726
Db	487	GAAAGCATTAAAGAAAACATTTCGTATATACAAACCTTTNTTTNTTATCTACTGATTT	546
Oy	727	TTTAAAGCTTACCCGTGACACTTTATCTAGTTGAATTACTCGATGTAC	777
Db	547	TGACAAAATTAATACCTTTTAAAGTTTAAACCCGAGCACTTNTTCTCTTGTCC	597

```

RESULT 2
US-09-385-982-133
Sequence 133, Application US/09385982
Patent No. 6262334
GENERAL INFORMATION:
APPLICANT: ENEGE, WILSON O., ET AL.
TITLE OF INVENTION: NOVEL HUMAN GENES AND GENE EXPRESSION
FILE OF INVENTION: PRODUCTS: 11
FILE REFERENCE: CCDNA-260XX
CURRENT APPLICATION NUMBER: US/09/385,982
EARLIER FILING DATE: 1999-08-30
EARLIER APPLICATION NUMBER: 09/328,111
EARLIER FILING DATE: 1999-06-08
EARLIER APPLICATION NUMBER: 60/117,393
EARLIER FILING DATE: 1999-01-27
EARLIER APPLICATION NUMBER: 60/099,639
EARLIER FILING DATE: 1998-08-31
NUMBER OF SEQ. ID NOS: 544
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 133
LENGTH: 588
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc-feature
LOCATION: (1)...(388)
OTHER INFORMATION: n = A,T,C or G
US-09-385-982-133

```

Query Match	Similarity	5.4%	Score 68	DB 3	length 588
Best Local	Similarity	56.1%	Pred. No. 5.4e-09		
Matches	Conservative	0	Mismatches	113	Indels
				10	Gaps
QY	361	TATCTAAATAGTTTAAAAAGCTGTAAAGCCGAGCAGCAGTTCGTACACCCAGTTACTA	420		
Db	189	TATNNNAATATPTNNNNNAANNTATTAANCTGCANCAANNTGATTTTNACACCTANTTACTA	248		
QY	421	GAAACGGAAGGAGGAGCTAGTACGCTGATGTAAGGAAGTGGAACAGGAACGACTTC	480		
Db	249	GAAACCTAANGAAGAGCACTNATTTAGCT-----CTGAATNAANTNACATGGNAACCTTT	302		
QY	481	TACTATCTACCAAAAAAATCTCCGAATGCATTTATCAGAAAGCTTTATATAGTACAGGTGAC	540		
Db	303	TACTATCTCTCNAANNAACCTTC-TCCTCAANTATNNNAAGATTATTATNACA--ANGNG	360		
QY	541	ACATATTGCTGTTTAGAAGGGGGTCTTAAAGAAAAAGCACTTGCTAAAGTTAGCAACTGTG	600		
Db	361	GNNATCTCCTNATCATATANNNGGTTCTATTANANA--CCCGCTAANTNTGCGCACTTAC	418		
QY	601	AGGATGGCCATTTTAAATATATGAGACCAAGCCCACTGG	640		
Db	419	AGAACANCCAGCANTANANATNTTTCATGCCCCATTTGGG	458		

RESULT 3  
 US-08-232-463-14/C  
 Sequence 14, Application US/08232463  
 Patent No. 5670367  
 GENERAL INFORMATION:  
 APPLICANT: DORNER, F.  
 APPLICANT: SCHEIFLINGER, F.  
 APPLICANT: FALKNER, F. G.  
 TITLE OF INVENTION: RECOMBINANT FOXP2 VIRUS  
 NUMBER OF SEQUENCES: 52  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Foley & Lardner  
 STREET: 1800 Diagonal Road, Suite 500  
 CITY: Alexandria  
 STATE: VA  
 COUNTRY: USA  
 ZIP: 22313-0299  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: Patentin Release #1.0, Version #1.25  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/232,463  
 FILING DATE:  
 CLASSIFICATION: 435  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US/07/935,313  
 FILING DATE:  
 APPLICATION NUMBER: EP 91 114 300.6  
 FILING DATE: 26-AUG-1991  
 ATTORNEY/AGENT INFORMATION:  
 NAME: BENT, Stephen A.  
 REGISTRATION NUMBER: 29,768  
 REFERENCE/DOCKET NUMBER: 30472/114 IMMU  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (703)836-9300  
 TELEFAX: (703)683-4109  
 TELEX: 899149  
 INFORMATION FOR SEQ ID NO: 14:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 7218 base pairs  
 TYPE: nucleic acid  
 STRANDEDNESS: single  
 TOPOLOGY: linear  
 IMMEDIATE SOURCE:  
 CLONE: pTZypc-F15  
 US-08-232-463-14

[illegible]





```
EARLIER APPLICATION NUMBER: 60/048, 949
EARLIER FILING DATE: 1997-06-06
EARLIER APPLICATION NUMBER: 60/048, 974
EARLIER FILING DATE: 1997-06-06
EARLIER APPLICATION NUMBER: 60/048, 883
EARLIER FILING DATE: 1997-06-06
EARLIER APPLICATION NUMBER: 60/048, 897
EARLIER FILING DATE: 1997-06-06
EARLIER APPLICATION NUMBER: 60/048, 898
EARLIER FILING DATE: 1997-06-06
EARLIER APPLICATION NUMBER: 60/048, 962
EARLIER FILING DATE: 1997-06-06
EARLIER APPLICATION NUMBER: 60/048, 963
EARLIER FILING DATE: 1997-06-06
EARLIER APPLICATION NUMBER: 60/048, 877
EARLIER FILING DATE: 1997-06-06
EARLIER APPLICATION NUMBER: 60/048, 878
EARLIER FILING DATE: 1997-06-06
EARLIER APPLICATION NUMBER: 60/070, 923
EARLIER FILING DATE: 1997-12-18
EARLIER APPLICATION NUMBER: 60/092, 921
EARLIER FILING DATE: 1998-07-15
EARLIER APPLICATION NUMBER: 60/094, 657
EARLIER FILING DATE: 1998-07-30
NUMBER OF SEQ ID NOS: 1227
SOFTWARE: Patentln Ver. 2.0
SEQ ID NO 103
LENGTH: 2218
TYPE: DNA
ORGANISM: Homo sapiens
US-09-205-258-103
```

```
Query Match 3.88; Score 47.8; DB 4; Length 2218;
Best Local Similarity 57.84; Pred. No. 0.0027;
Matches 85; Conservative 0; Mismatches 62; Indels 0; Gaps 0;
```

```
QY 1 TTTTCTTTTCTTTTCAAGTCCAAAGACATTTTCTTTTATGATCGAAG 60
    |||||||
DB 2207 TTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTT 2148
    |||||||
QY 61 ATTATTAAGTCAATACGCAAAACATCTGCTACTGCTTGCAGCAAAAGATCAATGTA 120
    |||||
DB 2147 GTGCTTTTATTTACTACTCAAAAAGAAACAAAGATGATGCTATCAAAAAGACATTTTAC 2088
    |||||
QY 121 AAGACTCCACATTTCTGCACGTCTCA 147
    |||||
DB 2087 AAGCTAAGATGATCAATGCTTCA 2061
    |||||
```

```
RESULT 6
US-09-205-258-233/C
Sequence 232, Application US/09205258
Patent No. 6525174
GENERAL INFORMATION:
APPLICANT: YOUNG et al.
TITLE OF INVENTION: 207 Human Secreted Proteins
FILE REFERENCE: P2007P1
CURRENT APPLICATION NUMBER: US/09/205, 258
EARLIER FILING DATE: 1998-12-04
EARLIER APPLICATION NUMBER: PCT/US98/11422
EARLIER FILING DATE: 1998-06-04
EARLIER APPLICATION NUMBER: 60/048, 885
EARLIER FILING DATE: 1997-06-06
EARLIER APPLICATION NUMBER: 60/049, 375
EARLIER FILING DATE: 1997-06-06
EARLIER APPLICATION NUMBER: 60/048, 881
EARLIER FILING DATE: 1997-06-06
EARLIER APPLICATION NUMBER: 60/048, 880
EARLIER FILING DATE: 1997-06-06
EARLIER APPLICATION NUMBER: 60/048, 896
EARLIER FILING DATE: 1997-06-06
EARLIER APPLICATION NUMBER: 60/049, 020
EARLIER FILING DATE: 1997-06-06
```

```
EARLIER APPLICATION NUMBER: 60/048, 876
EARLIER FILING DATE: 1997-06-06
EARLIER APPLICATION NUMBER: 60/048, 895
EARLIER FILING DATE: 1997-06-06
EARLIER APPLICATION NUMBER: 60/048, 884
EARLIER FILING DATE: 1997-06-06
EARLIER APPLICATION NUMBER: 60/048, 894
EARLIER FILING DATE: 1997-06-06
EARLIER APPLICATION NUMBER: 60/048, 971
EARLIER FILING DATE: 1997-06-06
EARLIER APPLICATION NUMBER: 60/048, 964
EARLIER FILING DATE: 1997-06-06
EARLIER APPLICATION NUMBER: 60/048, 882
EARLIER FILING DATE: 1997-06-06
EARLIER APPLICATION NUMBER: 60/048, 899
EARLIER FILING DATE: 1997-06-06
EARLIER APPLICATION NUMBER: 60/048, 893
EARLIER FILING DATE: 1997-06-06
EARLIER APPLICATION NUMBER: 60/048, 900
EARLIER FILING DATE: 1997-06-06
EARLIER APPLICATION NUMBER: 60/048, 901
EARLIER FILING DATE: 1997-06-06
EARLIER APPLICATION NUMBER: 60/048, 892
EARLIER FILING DATE: 1997-06-06
EARLIER APPLICATION NUMBER: 60/048, 915
EARLIER FILING DATE: 1997-06-06
EARLIER APPLICATION NUMBER: 60/049, 019
EARLIER FILING DATE: 1997-06-06
EARLIER APPLICATION NUMBER: 60/048, 970
EARLIER FILING DATE: 1997-06-06
EARLIER APPLICATION NUMBER: 60/048, 972
EARLIER FILING DATE: 1997-06-06
EARLIER APPLICATION NUMBER: 60/048, 916
EARLIER FILING DATE: 1997-06-06
EARLIER APPLICATION NUMBER: 60/049, 373
EARLIER FILING DATE: 1997-06-06
EARLIER APPLICATION NUMBER: 60/048, 875
EARLIER FILING DATE: 1997-06-06
EARLIER APPLICATION NUMBER: 60/049, 374
EARLIER FILING DATE: 1997-06-06
EARLIER APPLICATION NUMBER: 60/048, 917
EARLIER FILING DATE: 1997-06-06
EARLIER APPLICATION NUMBER: 60/048, 949
EARLIER FILING DATE: 1997-06-06
EARLIER APPLICATION NUMBER: 60/048, 974
EARLIER FILING DATE: 1997-06-06
EARLIER APPLICATION NUMBER: 60/048, 883
EARLIER FILING DATE: 1997-06-06
EARLIER APPLICATION NUMBER: 60/048, 897
EARLIER FILING DATE: 1997-06-06
EARLIER APPLICATION NUMBER: 60/048, 898
EARLIER FILING DATE: 1997-06-06
EARLIER APPLICATION NUMBER: 60/048, 962
EARLIER FILING DATE: 1997-06-06
EARLIER APPLICATION NUMBER: 60/048, 963
EARLIER FILING DATE: 1997-06-06
EARLIER APPLICATION NUMBER: 60/048, 877
EARLIER FILING DATE: 1997-06-06
EARLIER APPLICATION NUMBER: 60/048, 878
EARLIER FILING DATE: 1997-06-06
EARLIER APPLICATION NUMBER: 60/070, 923
EARLIER FILING DATE: 1997-12-18
EARLIER APPLICATION NUMBER: 60/092, 921
EARLIER FILING DATE: 1998-07-15
EARLIER APPLICATION NUMBER: 60/094, 657
EARLIER FILING DATE: 1998-07-30
NUMBER OF SEQ ID NOS: 1227
SOFTWARE: Patentln Ver. 2.0
SEQ ID NO 232
LENGTH: 760
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
```

```

: NAME/KEY: SITE
: LOCATION: (438)
: OTHER INFORMATION: n equals a,t,g, or c
:
: FEATURE:
: NAME/KEY: SITE
: LOCATION: (741)
: OTHER INFORMATION: n equals a,t,g, or c
NS-09-203-238-232

```

Query Match	3.7%	Score 46.8;	DB 4;	Length 760;
Best Local Similarity	57.1%;	Pred. No. 0.0032;		
Matches 84; Conservative	0;	Mismatches 63;	Indels 0;	Gaps 0;

[illegible]

OY 61ATTTAAAGTCATACATGCAAAAATCTGTACTGCATTAGCAAAAGATCATGTAA 120  
| | | | | | | | | | | | | | | | | |  
Db 688 GCGTTTATTATTAACCTAACAAAAAAGAACAAGATGATGGTATCAAAAAGACAATTTAC 629

```

qy      121 AAACACTCCACAATTCTGCACACTGTCA 147
        |||| | | | | | | | |
Db      628 AAACCTAGAATAGTAACATAGCTTTCA 602

```

RESULT 7  
US-09-734-675-3/c  
Sequence 3 Annotation: NC\_00734675

```

; GENERAL INFORMATION:
; APPLICANT: WEBSTER, Marion et al
; TITLE OF INVENTION: ISOLATED HUMAN

```

1  
 2  
 3  
 4  
 5  
 6  
 7  
 8  
 9  
 10  
 11  
 12  
 13  
 14  
 15  
 16  
 17  
 18  
 19  
 20  
 21  
 22  
 23  
 24  
 25  
 26  
 27  
 28  
 29  
 30  
 31  
 32  
 33  
 34  
 35  
 36  
 37  
 38  
 39  
 40  
 41  
 42  
 43  
 44  
 45  
 46  
 47  
 48  
 49  
 50  
 51  
 52  
 53  
 54  
 55  
 56  
 57  
 58  
 59  
 60  
 61  
 62  
 63  
 64  
 65  
 66  
 67  
 68  
 69  
 70  
 71  
 72  
 73  
 74  
 75  
 76  
 77  
 78  
 79  
 80  
 81  
 82  
 83  
 84  
 85  
 86  
 87  
 88  
 89  
 90  
 91  
 92  
 93  
 94  
 95  
 96  
 97  
 98  
 99  
 100  
 101  
 102  
 103  
 104  
 105  
 106  
 107  
 108  
 109  
 110  
 111  
 112  
 113  
 114  
 115  
 116  
 117  
 118  
 119  
 120  
 121  
 122  
 123  
 124  
 125  
 126  
 127  
 128  
 129  
 130  
 131  
 132  
 133  
 134  
 135  
 136  
 137  
 138  
 139  
 140  
 141  
 142  
 143  
 144  
 145  
 146  
 147  
 148  
 149  
 150  
 151  
 152  
 153  
 154  
 155  
 156  
 157  
 158  
 159  
 160  
 161  
 162  
 163  
 164  
 165  
 166  
 167  
 168  
 169  
 170  
 171  
 172  
 173  
 174  
 175  
 176  
 177  
 178  
 179  
 180  
 181  
 182  
 183  
 184  
 185  
 186  
 187  
 188  
 189  
 190  
 191  
 192  
 193  
 194  
 195  
 196  
 197  
 198  
 199  
 200  
 201  
 202  
 203  
 204  
 205  
 206  
 207  
 208  
 209  
 210  
 211  
 212  
 213  
 214  
 215  
 216  
 217  
 218  
 219  
 220  
 221  
 222  
 223  
 224  
 225  
 226  
 227  
 228  
 229  
 230  
 231  
 232  
 233  
 234  
 235  
 236  
 237  
 238  
 239  
 240  
 241  
 242  
 243  
 244  
 245  
 246  
 247  
 248  
 249  
 250  
 251  
 252  
 253  
 254  
 255  
 256  
 257  
 258  
 259  
 260  
 261  
 262  
 263  
 264  
 265  
 266  
 267  
 268  
 269  
 270  
 271  
 272  
 273  
 274  
 275  
 276  
 277  
 278  
 279  
 280  
 281  
 282  
 283  
 284  
 285  
 286  
 287  
 288  
 289  
 290  
 291  
 292  
 293  
 294  
 295  
 296  
 297  
 298  
 299  
 300  
 301  
 302  
 303  
 304  
 305  
 306  
 307  
 308  
 309  
 310  
 311  
 312  
 313  
 314  
 315  
 316  
 317  
 318  
 319  
 320  
 321  
 322  
 323  
 324  
 325  
 326  
 327  
 328  
 329  
 330  
 331  
 332  
 333  
 334  
 335  
 336  
 337  
 338  
 339  
 340  
 341  
 342  
 343  
 344  
 345  
 346  
 347  
 348  
 349  
 350  
 351  
 352  
 353  
 354  
 355  
 356  
 357  
 358  
 359  
 360  
 361  
 362  
 363  
 364  
 365  
 366  
 367  
 368  
 369  
 370  
 371  
 372  
 373  
 374  
 375  
 376  
 377  
 378  
 379  
 380  
 381  
 382  
 383  
 384  
 385  
 386  
 387  
 388  
 389  
 390  
 391  
 392  
 393  
 394  
 395  
 396  
 397  
 398  
 399  
 400  
 401  
 402  
 403  
 404  
 405  
 406  
 407  
 408  
 409  
 410  
 411  
 412  
 413  
 414  
 415  
 416  
 417  
 418  
 419  
 420  
 421  
 422  
 423  
 424  
 425  
 426  
 427  
 428  
 429  
 430  
 431  
 432  
 433  
 434  
 435  
 436  
 437  
 438  
 439  
 440  
 441  
 442  
 443  
 444  
 445  
 446  
 447  
 448  
 449  
 450  
 451  
 452  
 453  
 454  
 455  
 456  
 457  
 458  
 459  
 460  
 461  
 462  
 463  
 464  
 465  
 466  
 467  
 468  
 469  
 470  
 471  
 472  
 473  
 474  
 475  
 476  
 477  
 478  
 479  
 480  
 481  
 482  
 483  
 484  
 485  
 486  
 487  
 488  
 489  
 490  
 491  
 492  
 493  
 494  
 495  
 496  
 497  
 498  
 499  
 500  
 501  
 502  
 503  
 504  
 505  
 506  
 507  
 508  
 509  
 510  
 511  
 512  
 513  
 514  
 515  
 516  
 517  
 518  
 519  
 520  
 521  
 522  
 523  
 524  
 525

```

; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSeq for Windows Version 4.0.0
; SEQ ID NO 3

```

```

; TYPE: DNA
; ORGANISM: Human
US-09-734-675-3

```

Query Match	3.6%;	Score 45.6;	DB 4;	Length 38844;
Best Local Similarity	55.8%;	Pred. NO. 0.036;		
Matches 87; Conservative	0;	Mismatches 69;	Indels 0;	

[illegible]

0y    TTTATTAAGCATACATGC~~AA~~ACATACAGCTAACTAGCATTTAGCAAAAGATCAATGTAA 121  
 62    TTTATTAAAGTCATACATGC~~AA~~ACATACAGCTAACTAGCATTTAGCAAAAGATCAATGTAA 121  
 db    TTTATTAAAGTCATACATGC~~AA~~ACATACAGCTAACTAGCATTTAGCAAAAGATCAATGTAA 121  
 24548    TTTATTAAAGTCATACATGC~~AA~~ACATACAGCTAACTAGCATTTAGCAAAAGATCAATGTAA 121

Dy	I22	AACACTCCACAATTTCGCACCTGCAATGTGAAGAAAAA	157
Dd	24488	ACAGTTCCTAAATAGTAAAAATTCGCTTTTAAAAAGAA	24453

RESULT 8  
US-09-370-253-1/c

;; Patent No. 6165792  
;; GENERAL INFORMATION:  
: APPLICANT: Allen, Stephen M.

```

; APPLICANT: Thorpe, Catherine J.
; TITLE OF INVENTION: Amino Acid Transporters
; FILE REFERENCE: BB-1200

```

```

: CURRENT APPLICATION NUMBER: US/09/370,255
: CURRENT FILING DATE: 1999-08-09
: EARLIER APPLICATION NUMBER: 60/097,222
: EARLIER FILING DATE: August 20, 1998
: NUMBER OF SEQ ID NOS: 18
: SOFTWARE: Microsoft Office 97
: SEQ ID NO 1
: LENGTH: 1882
: TYPE: DNA
: ORGANISM: Hordeum vulgare
US-09-370-253-1

```

Query Match	3.6%	Score 45;	DB 3;	Length 1082;
Best Local Similarity	54.5%;	Pred. NO. 0.015;		
Matches 90;	Conservative 0;	Mismatches 75;	Indels 0;	Gaps 0;

**Oy** TTTTCTTTTTTTTTTTCAAGTCACAGACATTTTTTTTTTTTATGATCAGG 60  
|||||  
**Db** TTTTCTTTTTTTTTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTT 180

Oy	61 AATTATTAAGTCACATCATCAAAACATACTGCTA	CTGCATTAGCAAAGAATCATGTGA	120
Db	1800 TTTTITTTTTTTTTTGAAAGAAAAGAACCCATA	CATTGGCGAATPATCTATGCCAA	1743

QY 121 AAACACTCCACAATTCGTGCACTGTCAATTGA AAAAGTTTGTC 165  
||| | | | | | | | | |  
Db 1740 GAACAATCGCAATGTATGAAATCTGCACCTAATTTGATATTTCTC 1696

RESULT 9  
US-08-924-747-25/C

GENERAL INFORMATION:  
APPLICANT: MCGONIGLE, BRIAN  
ADDRESS: 6700 W. 12TH AVE.  
CITY: DENVER, CO 80202  
STATE: CO  
COUNTRY: USA  
DATE: 01/10/80  
TIME: 10:00  
PAGE: 1

```

; ; TITLE OF INVENTION: SOYBEAN GLUTATHIONE-S-TRANSFERASE
; ; TITLE OF INVENTION: ENZYMES
; ; NUMBER OF SEQUENCES: 32
; ; CORRESPONDENCE ADDRESS:

```

ADDRESSEE: E.I. DU PONT DE NEMOURS AND COMPANY  
STREET: 1007 MARKET STREET  
CITY: WILMINGTON  
STATE: DELAWARE

```

; COUNTRY: UNITED STATES OF AMERICA
; ZIP: 19898
; COMPUTER READABLE FORM:
; MEDIUM TYPE: DISKETTE 3 50 INCH

```

```

;
; COMPUTER: IBM PC COMPATIBLE
; OPERATING SYSTEM: MICROSOFT WORD FOR WINDOWS 95
; SOFTWARE: MICROSOFT WORD VERSION 7.0A
; CURRENT APPLICATION DATA:

```

; APPLICATION NUMBER: US/08/924,747  
 ;  
 ; FILING DATE:  
 ;  
 ; CLASSIFICATION: 435  
 ;  
 ; ATTORNEY/AGENT INFORMATION:  
 ;

NAME: FLOYD, LINDA AXAMEETHY  
REGISTRATION NUMBER: 33,692  
REFERENCE/DOCKET NUMBER: CL-1108  
TELECOMMUNICATION INFORMATION:

TELEPHONE: 302-892-8112  
TELEFAX: 302-773-0164  
INFORMATION FOR SEQ ID NO: 25:  
SEQUENCE CHARACTERISTICS.

```

; LENGTH: 991 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear

```

```

; MOLECULE TYPE: CDNA
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:

```

TISSUE TYPE: SOYBEAN



Search completed: August 26, 2003, 02:33:12  
Job time : 94.8689 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleole - nucleole search, using sw model

Run on: August 25, 2003, 22:04:33 ; Search time 309.402 Seconds  
(without alignments)  
9198.941 Million cell updates/sec

Title: US-09-717-321a-2

Perfect score: 1266  
Sequence: 1 ttttttttttttttttcaa.....gagcaagaatagggtatttt 1266

Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 1517243 seqs, 1124081882 residues

Total number of hits satisfying chosen parameters: 3034486

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Published\_Applications\_MA.\*

1: /cgn2\_6/ptodata/2/pubpna/US07\_PUBCOMB.seq.\*  
2: /cgn2\_6/ptodata/2/pubpna/PCT\_NEW\_PUB.seq.\*  
3: /cgn2\_6/ptodata/2/pubpna/US06\_NEW\_PUB.seq.\*  
4: /cgn2\_6/ptodata/2/pubpna/US06\_PUBCOMB.seq.\*  
5: /cgn2\_6/ptodata/2/pubpna/US07\_NEW\_PUB.seq.\*  
6: /cgn2\_6/ptodata/2/pubpna/PCTUS\_PUBCOMB.seq.\*  
7: /cgn2\_6/ptodata/2/pubpna/US08\_NEW\_PUB.seq.\*  
8: /cgn2\_6/ptodata/2/pubpna/US08\_PUBCOMB.seq.\*  
9: /cgn2\_6/ptodata/2/pubpna/US09A\_PUBCOMB.seq.\*  
10: /cgn2\_6/ptodata/2/pubpna/US09C\_PUBCOMB.seq.\*  
11: /cgn2\_6/ptodata/2/pubpna/US09C\_NEW\_PUB.seq.\*  
12: /cgn2\_6/ptodata/2/pubpna/US09A\_PUBCOMB.seq.\*  
13: /cgn2\_6/ptodata/2/pubpna/US10A\_PUBCOMB.seq.\*  
14: /cgn2\_6/ptodata/2/pubpna/US10B\_PUBCOMB.seq.\*  
15: /cgn2\_6/ptodata/2/pubpna/US10\_NEW\_PUB.seq.\*  
16: /cgn2\_6/ptodata/2/pubpna/US60\_NEW\_PUB.seq.\*  
17: /cgn2\_6/ptodata/2/pubpna/US60\_PUBCOMB.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	458.8	36.2	2941	14	US-10-198-846-9874 Sequence 9874, App
C 2	447.6	35.4	2051	9	US-09-925-302-255 Sequence 255, App
C 3	430.6	34.0	1644	14	US-10-153-668-48 Sequence 48, Appl
C 4	430.6	34.0	1646	14	US-10-153-668-46 Sequence 46, Appl
C 5	420.2	33.2	1232	10	US-09-880-107-1632 Sequence 1632, App
C 6	284.2	22.4	2313	14	US-10-198-846-12919 Sequence 12919, A
C 7	195.2	15.4	594	14	US-10-066-543-2864 Sequence 2864, App
C 8	193.2	15.3	602	10	US-09-878-178-1654 Sequence 1654, App
C 9	193.2	15.3	602	13	US-10-046-935-1654 Sequence 1654, App
C 10	193.2	15.3	602	14	US-10-146-502-1654 Sequence 1654, App
C 11	187.8	14.8	549	14	US-10-066-543-2878 Sequence 2878, App
C 12	187.8	14.8	567	14	US-10-066-543-2964 Sequence 2964, App
C 13	187.2	14.8	605	11	US-09-871-161-483 Sequence 483, App
C 14	185.4	14.6	477	14	US-10-066-543-3022 Sequence 3022, App
C 15	172.2	13.6	432	10	US-09-960-352-7612 Sequence 7612, App
C 16	164	13.0	2167	12	US-10-101-510-500 Sequence 500, App

17	163.6	12.9	353	10	US-09-960-352-11432	Sequence 11432, A
C 18	147.4	11.6	409	14	US-10-066-543-13	Sequence 13, Appl
C 19	126.4	10.0	285	9	US-09-294-0938-506	Sequence 506, App
C 20	123.2	9.7	416	10	US-09-960-352-8781	Sequence 8781, App
C 21	116.4	9.2	3740	11	US-09-764-891-9984	Sequence 9984, App
C 22	108.8	8.6	293	14	US-10-066-543-3077	Sequence 3077, App
C 23	98.2	7.8	581	10	US-09-988-598-972	Sequence 972, App
C 24	95.4	7.5	201	10	US-09-960-352-4357	Sequence 4357, App
C 25	91.6	7.2	565	10	US-09-998-598-296	Sequence 296, App
C 26	85.6	6.8	492	10	US-09-878-178-1503	Sequence 1503, App
C 27	85.6	6.8	492	13	US-10-046-935-1503	Sequence 1503, App
C 28	85.6	6.8	492	14	US-10-146-502-1503	Sequence 1503, App
C 29	82.2	6.5	398	11	US-09-918-995-34569	Sequence 34569, A
C 30	68	5.4	588	11	US-09-871-161-133	Sequence 133, App
C 31	62.2	4.9	426	10	US-09-960-352-7342	Sequence 7342, App
C 32	60	4.7	404	10	US-09-960-352-14206	Sequence 14206, A
C 33	57.6	4.5	1651	14	US-10-198-846-13019	Sequence 13019, A
C 34	52.4	4.1	162	10	US-09-920-300A-1549	Sequence 1549, App
C 35	52.4	4.1	162	13	US-10-033-528-1549	Sequence 1549, App
C 36	52.2	4.1	424	10	US-09-960-352-11218	Sequence 11218, A
C 37	49.4	3.9	469	11	US-09-918-995-13017	Sequence 13017, A
C 38	48.2	3.8	277	10	US-09-960-352-12673	Sequence 12673, A
C 39	48	3.8	640681	10	US-09-790-988-1	Sequence 1, Appl1
C 40	47.8	3.8	2218	14	US-10-023-282-103	Sequence 103, App
C 41	46.8	3.7	408	10	US-09-960-352-6263	Sequence 6263, App
C 42	46.8	3.7	760	14	US-10-023-282-232	Sequence 232, App
C 43	46.4	3.7	15261	11	US-09-764-891-10186	Sequence 10186, A
C 44	45.8	3.6	1008	9	US-09-780-641-1	Sequence 1, Appl1
C 45	45.8	3.6	3770	9	US-09-925-302-346	Sequence 346, App

## ALIGNMENTS

RESULT 1	US-10-198-846-9874/c	Application US/10198846
Sequence 9874,	Publication No. US2003009974A1	
GENERAL INFORMATION:		
APPLICANT:	Lillie, James	
APPLICANT:	Xu, Yongyao	
APPLICANT:	Wang, Youzhen	
APPLICANT:	Stelmann, Kathleen	
TITLE OF INVENTION:	NOVEL GENES, COMPOSITIONS, KITS, AND METHODS	
TITLE OF INVENTION:	FOR IDENTIFICATION, ASSESSMENT, PREVENTION, AND	
TITLE OF INVENTION:	THERAPY OF BREAST CANCER	
FILE REFERENCE:	MRI-049	
CURRENT APPLICATION NUMBER:	US/10/198,846	
CURRENT FILING DATE:	2002-07-18	
PRIOR APPLICATION NUMBER:	60/306,220	
PRIOR FILING DATE:	2001-07-18	
NUMBER OF SEQ ID NOS:	14064	
SOFTWARE:	FastSeq for Windows Version 4.0	
SEQ ID NO 9874		
LENGTH:	2941	
TYPE:	DNA	
ORGANISM:	Homo sapiens	
US-10-198-846-9874		
Query Match	36.2%;	Score 458.8; DB 14; Length 2941;
Best Local Similarity	68.0%;	Pred. No. 2,6e-115;
Matches 895;	Conservative	0; Mismatches 337; Indels 84; Gaps 15;
QY	1	TTTTTTTTTTTTTTTTCAGTCCAAAGACATTTTTTTTTTTTTTTTATGATCAGG 60
DB	2481	TTTATATCAAGG 2422
QY	61	ATTATTAAGTATCATGCAAAACATCTGCTAATGATAGCAAAAGATCAATGTGA 120
DB	2421	ATTATTAAGTATCATGCAAAACATCTGCTAATGATAGCAAAAGATCAATGTGA 2362
QY	121	AAACTCCACATCTGCAACTGTCAATGAAAAAGTTGTCTAGTGTGCAAGGC 180

236	AAACGTCACAAATTCGCAACGTGCAATTAAAAAAATTCGTGTAGTGAGGCTGAAGGCT	2302
QY	181 CCAGCACTGTGTCTTCCAGTAGTGGTTAGTGTGTACAGAAAGCGGTTAGCACTAGCC--	238
Db	2301 CCCAGCGGTGTATTCCTCCAGTGGTTAAAGTTGTACAGAAACATGTGAGCACTAGCAAG	2242
QY	239 TTGCACAAACCTCACAAACCCAAAG-----TACGGAAGCATG	277
Db	2241 TTTCACAGAACTCAGACAGCCCAAGAAACATCATATAGGCAAAAGGACTACAGAGCGGTG	2182
QY	278 TGTCCCGGTGGGTAGGCTTAGAGGGGGCGGCATCAATCACATGACAGTGTGGTACTCT	337
Db	2181 TGTCCGGGTGGCGCAGGTAAAGA--GGGTCACTATTGGTCAAGTACAGTGTGCGGTATCT	2123
QY	338 GGCAGACAGAGGAG-----TTTCAGAAATATCTAAAAATGGTTAAAAA	380
Db	2122 GGCAGACAGAGTGTATGTATAGAAAGTCTTACGTATTATAGCAATTTATCTAAAAATTTTAAAAA	20633
QY	381 CTGTAAAGCCGACAGAGTGAATTTCTACACCAGTTACTAGAAAAAGAAAGGACGCTA	440
Db	2062 CTATTAAGAGCTGCACACATGATTTTTACACTAGTTACTAGAAAACTTAAGAAAGCACTT	20033
QY	441 GTCAGCTGAGTAAAGGAAGGTGAAAAACGAAACCCACTTCTACATCTACCAAAAAATC	500
Db	2002 ATTAGCTGTGATA-----AAGTACATGGAACACACTTTACGTATCTGACAAAAAAACC	1948
QY	501 TCCGAATGCAATTATCAGAAAGATCTTATAGTACAGAGTCAGACATATTGCTGTTAAGAG	560
Db	1947 TTCTATATGCAATTATCAGAAAGATTTTATATATACAAG--GAGGCATATTGCTCAGTCAGAG	1889
QY	561 GGGGTCCCTAAAGAAAGACACTTGCTAGTTAGCACTGTGAGATGGCCACTTTAAATAT	620
Db	1888 GGGGTCTATTAAGAAAGACCTTACTTAAGTTCGCGCTAACAGAAACACACTTTAAAGAT	1829
QY	621 GGACTCAAGCGCCCATCTGGGAGAGGACAGACAGAGGGAAGGGGGGCTCAAGAGACACT	680
Db	1828 GAATTTAAATGGCCCAATTTGGGAGAGCATGCGAGGTGAAGAAAGAAAGGCTTTAAGA	1768
QY	681 GATTAAGATCGCCACTTTGTCTACTACTGTGTTGACAGAAATTAACCGTTAAAAAGCTTTAC	740
Db	1768 AACATTTCTCATTAATACCAACCTTTCTTCAATCATCTACGATTTGACGAATTAATAC	1709
QY	741 CCGAGACACTTTTATTCAGTTGAA---TTACTCAGTACAAATAGTGTAAATTAATCT	797
Db	1708 CTTTATAGAGTTTATACCGGTACACTTTTCAATCTGTGACAAATGAGTGA---AATCT	16533
QY	798 CTACTTCATATTA--GTCAAAATACTGTCTCTCTCTTGAATGAGTGTGCTTTACACAC	856
Db	1652 CCACCTGCTATTTGTCAAAATACTGTCTTGTGCTTTGATCA-----CACACAC	16033
QY	857 TTCACCCGACACCCACAGACTTAGGAACAGAAATACTTGTTAGAGGCAACAGAGACCA	916
Db	1602 CCCACCCGGCACCCACAGCTA--AACAGAAATCTTCAATTAGAGGAATA-----GCA	1551
QY	917 GAGTTCCTTTAAAGCCGTGACAAAGCCGCTAGCTGGAATTTAGAGAACTCACTATGA	976
Db	1550 GTTCTGTCAAAATCTCCGCAAAAGCTGTGAGAAATCCGCTATGAAATCACAAGACT	1491
QY	977 ATCAAAAGCAGAGCTGTTCACACCATCTGTGACGCTACGTACGAAAGTTACGTATAGCA	1038
Db	1490 GATCCAAAGACTGAGCTGTACTACGCTCACTCCATTCACGAAATGTTATGTGGGAACA	1431
QY	1037 TGGGCTGAATAGTTTACAGTGCCTTACATATGGAGAGGTCATTTAAAGAGGGCTGTGCTG	1096
Db	1430 CGTGTCTTAACCTACGTGTAGTTCAATGGCAACGCTTTCATTTGGGAGGCTGTCTGCT	1371
QY	1097 TCACAGGCTG-----GAGCTACGGAGGCTCTGACC--CTGAGCCAGAAAGCTG	1148
Db	1370 TTACGCACTGTAGAACTACATAGGAGAGCAAGTGTCTCACCTCTCAATGAGAAAGCTA	1311
QY	1149 CAGCTCTTTAAAGCAAAAGTCTGCACACACTTATAGTGTCTTACGTGTCTCAGCACAAAC	1207
Db	1310 CGCTCTTTTCAAAAGCAAGGCTCTTTGCAAAGTTCAAGTGTGCGGTGTCTGTGGCACAA	1251

QY	1208	-GCAACTTGTTCACAGGATATTTGGCAATCTTAATCTGACCAAGATTAGGG	1261
Db	1250	ATGCAGTGTAAGTTCAGAAAGGATTTTGGCAACTCTTAATCTGACCAAGATTAGGGG	1195
RESULT 2			
US-09-925-302-255/c			
Sequence 255, Application US/09925302			
Patent No. US20020044941A1			
GENERAL INFORMATION:			
APPLICANT: Rosen et al.			
TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies			
FILE REFERENCE: PA104			
CURRENT APPLICATION NUMBER: US/09/925,302			
CURRENT FILING DATE: 2001-08-10			
PRIOR APPLICATION NUMBER: PCT/US00/05918			
PRIOR FILING DATE: 2000-03-08			
PRIOR APPLICATION NUMBER: 60/124,270			
PRIOR FILING DATE: 1999-03-12			
NUMBER OF SEQ ID NOS: 896			
SOFTWARE: PatentIn Ver. 2.0			
SEQ ID NO 255			
LENGTH: 2051			
TYPE: DNA			
ORGANISM: Homo sapiens			
FEATURE:			
NAME/KEY: misc feature			
LOCATION: (2)			
OTHER INFORMATION: n equals a,t,g, or c			
NAME/KEY: misc feature			
LOCATION: (50)			
OTHER INFORMATION: n equals a,t,g, or c			
NAME/KEY: misc feature			
LOCATION: (68)			
OTHER INFORMATION: n equals a,t,g, or c			
NAME/KEY: misc feature			
LOCATION: (2027)			
OTHER INFORMATION: n equals a,t,g, or c			
NAME/KEY: misc feature			
LOCATION: (2046)			
OTHER INFORMATION: n equals a,t,g, or c			
US-09-925-302-255			
Query Match 35.4%; Score 447.6; DB 9; Length 2051;			
Best Local Similarity 68.2%; Pred. No. 2.5e-112;			
Matches 876; Conservative 0; Mismatches 324; Indels 84; Gaps 15;			
QY	33	TTTTTTTTTTTTTTTATGATTCAGGATTTATTAAGTCATACATGCAAAACATATCTGC	92
Db	2009	TTTTTTTTTTTTTTTTTATGATTCAGGATTTATTAAGTCATACATGCAAAACATATATGC	1950
QY	93	TAACTGCATTAGCAAAAGATCAATGTAAAAACATCCCAATTTCTGCACTGTCAATTGA	152
Db	1949	TAAATGCAATTAGCAAAAGATCAATGTAAAAACATCCCAATTTCTGCACTGTCAATTGA	1890
QY	153	AAAAAGTTGTTAGTGTGTGGAAGGCCCAACACTGCTTCTGGCAGTAGTAGTTAGTT	212
Db	1889	AAAAATTTGT	1830
QY	213	GTACAGAAAGGCGTTAGCACTAGCGC--TTGACAGAAACCTCAGACAGCCCAAGG----	264
Db	1829	GTACAGAAACATCTGTCAGCACTAGCACTAGCACTAGCACTAGCACTAGCACTAGCACT	1770
QY	265	-----TACCGAAGCATGTGTCCGCGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT	309
Db	1769	AATAGCCAAAGCAGCTACAGAGAGCGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT	1711
QY	310	ATCAATCAATGACAGAGCTGT	352
Db	1710	ATTGCTCAAGTGACAGCTGT	1651
QY	353	TTTCAGATATCTAAATAGTTTAAATTAAGCCGACAGCAGTATTTCTACACCC	412



```
Db      1650 TTAAGAAATATCTAAATATTTTAAAAAAGCTCAACACATGATTTTACACT 1591
Qy      413 ACTTACTGAAAAGGAGGAGCACTACGAGTAAGAAAGGAAAGCAACAGAA 472
Db      1590 AGTTCTGAAAAGGAGGAGCACTTATAGCTGTGAAT- - - - -AAGTACATGGA 1536
Qy      473 GCGACTTCTACTATCTACCAAAAAAATCCGAATGCAATATCAGAAAGATCTTAA 532
Db      1535 AGCACTTTTACTATCGAACAAAAAAACCTTCAATGCAATATCAGAAAGATTTTAA 1476
Qy      533 CAGGTACAGATATTTGCTGTTAAGAGGGGCTCTAAGAAAGCACTTGTAAATAG 592
Db      1475 CAAG-GAGGCAATATGCTCAGTACAGAGGGGTTTATAGAAAGCACTTAAATAG 1417
Qy      593 CAAGTGTGAGATGGGAGGTTTAAATATGAGTCAAGCCCACTGGGAGGAGCAGA 652
Db      1416 GAGTAAACAGAACACAGTTTAAAGATGAATTTAAATGCCCAATTTGGGAGGAGTGA 1357
Qy      653 GGGGGAAGGGGGCTCAAGAGAGACATGATAGATCGGCCATTTGTCTACTGTG 712
Db      1356 GGTATAGAGAAAGGAAAGCTTAAAGAAACATTTCTGATATATACAACTTCTTCA 1297
Qy      713 ACAGAAATTAACCGTTAAAAAGCTTTACCGGTGACACTTTTATGAGTTGAA- - -TACT 769
Db      1296 TCATCTACTGCAATTTGACAGAAATTAACCTTTTAAAGTTTAAACCGGTGACACTTCAAT 1237
Qy      770 CCATGTACAAATGATGATTAATTAATCTACTACTATA-TTGTCAAAATATGCTGTC 828
Db      1236 CTTTATCAATATGATGTA- - - - -AATCTCACTGATGATTTGTCAAAATATGCTGTC 1181
Qy      829 TCCCTTGAAGAGCTGTGTTTCAACACTCCACCAGCAGCAGCCAGCTAGAGACAGA 888
Db      1180 TCCTTTGATCA- - - - - -CACACACCCAGCCGCGGACAGCCAGAGTA- - -MACAGA 1133
Qy      889 TACTTCTGTTAAGGCAACAGAGAGCCAGAGTTCTGTTCAAGCTCAGAAAGCCGTTCA 948
Db      1132 TTTCTTCAATTAAGGAAATA- - - - -GCAATTCTGTTCAAAAATCTCGCAAAAGCTGTCA 1079
Qy      949 GCTGTATTTTGAAGACACTACATGAATCAAAAGAGAGAGCTTTACACCACATCGTA 1008
Db      1078 GAAAACCTGCTATGAAATACAAAGACATGATCCAAAGAGCTGACTACCTCACTCC 1019
Qy      1009 GGTACAGTACAAAGTATGATAGATAGAGATGGCTGATATGATACAGTGTGCTTACATGAC 1068
Db      1018 ATTACAGTACAAATGTTATGTGGGAACAGTGTCTGTACATCACTGATGAGTTCAATGAC 959
Qy      1069 AGCGTCTATTAAAGAGGCTGTGCTGTGTGTACACAGGCTCG- - - - -GAGCTACGGGAG 1121
Db      958 AACGCTTCAATTCGGGAGGCTGTGCTTACGCACTGAGAACTACATAGAGAGAGCAG 899
Qy      1122 GGTGTGACAC- - -CCTGAGGCCAAGAGCTGAGTCTTAAAGGACAAAGTCTCAACAGC 1180
Db      898 TGTCTGACACTCTCTAATGCAAGCAAGCTACGCTTCTCAAGACAGAAAGTCTTTCAGAA 839
Qy      1181 TTAGTGTCTTACGTTCTCAGACAAAC- - -GCACTATGATTCAGAAAGTATTTGGCAAT 1237
Db      838 TTAGTGTCTTACGTTCTCAGACAAAC- - -GCACTATGATTCAGAAAGTATTTGGCAAT 779
Qy      1238 TCTTAAATCTGAGCAAGAAATAGGGG 1261
Db      778 TCTTAAATCTGAGCAAGAAATAGGGG 755
```

## RESULT 3

```
US-10-153-668-48/C
; Sequence 48, Application US/10153668
; Publication No. US20030092616A1
; GENERAL INFORMATION:
; APPLICANT: HONDA, Goichi
; APPLICANT: MATSUDA, Akio
; APPLICANT: MURAMATSU, Shuji
; APPLICANT: ISHIZAWA, Kenya
```

```
; TITLE OF INVENTION: STAT6 Activating Gene
; FILE REFERENCE: 1254-0207P
; CURRENT APPLICATION NUMBER: US/10/153,668
; PRIOR FILING DATE: 2002-05-24
; PRIOR APPLICATION NUMBER: US 60/293,172
; PRIOR FILING DATE: 2001-05-25
; PRIOR APPLICATION NUMBER: US 60/316,031
; PRIOR FILING DATE: 2001-08-31
; PRIOR APPLICATION NUMBER: US 60/328,403
; PRIOR FILING DATE: 2001-10-12
; PRIOR APPLICATION NUMBER: JP 2001-157043
; PRIOR FILING DATE: 2001-05-25
; PRIOR APPLICATION NUMBER: JP 2001-260681
; PRIOR FILING DATE: 2001-08-30
; PRIOR APPLICATION NUMBER: JP 2001-313175
; PRIOR FILING DATE: 2001-10-10
; NUMBER OF SEQ ID NOS: 488
; SOFTWARE: PatentIn Ver. 2.0
; LENGTH: 1644
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (109)..(447)
; US-10-153-668-48
```

Query Match 34.0%; Score 430.6; DB 14; Length 1644;

Best Local Similarity 67.8%; Pred. No. 1e-107;

Matches 859; Conservative 0; Mismatches 324; Indels 84; Gaps 15;

```
Qy      50 ATGATTCAGAGATTTATTAAGTCATACATGCAAAACATCTGCTAAGCTATGACAA 109
Db      1644 ATGATTCAGAGATTTATTAAGTCATACATGCAAAACATCTAATGCTATGACAA 1585
Qy      110 GATCAATGTAAAAACACTCCAAATTCGCAACTGTCAATTTGAAAAAAGTTGTTCTAG 169
Db      1584 GATCAATGTAAAAACACTCCAAATTCGCAACTGTCAATTTGAAAAAATTCGTTGAGT 1525
Qy      170 GGTCAAAAGGCCCAACACTGTGTTCTGTGCATGAGTAAAGTTGTACAGAGCGGTTAG 229
Db      1524 GGTCAAAAGGCCCAACACTGTGTTCTGTGCATGAGTAAAGTTGTACAGAGCGGTTAG 1465
Qy      230 CACTAGCCG- -TTGACAGAACCTCACAGACCCAAAG- - - - -TA 266
Db      1464 CACTAGCAAGTTTACAGAACTCTCACAGACCCAAAGAACATCATATAGCAAGGACTA 1405
Qy      267 CCGGAAGCATGTGTCGCGGTGGGTAGGTCTAGAGGGGCGGCATCAATCATGACAGT 326
Db      1404 CAGGAGCGGTGTGTCGCGGTGGGCGAGGTAAAGA- -GGTCAATATTGTCAGAGTCA 1346
Qy      327 GTTGTACTCTGGCAAGACAGTATG- - - - -TTTCAGAAATATCTAA 369
Db      1345 GTTGTACTCTGGCAAGACAGTATG- - - - -TTTCAGAAATATCTAA 1286
Qy      370 TAGTTTAAATCTGTAAGCCGACAGCAGTGTATCTACACCAGTATGACAAAGCA 429
Db      1285 TATTTTAAATCTATTAAGCTCAACACATGATTTTACACTGATTAAGTAAACATTA 1226
Qy      430 GGGAAAGCACTAGTCACTGAGTAAAGGAAAGTGAAGAAAGCAAGCACTTCTAGTCTA 489
Db      1225 GGAAGCACTATTAAGTCTGAAAT- - - - -AAGTAAATGGAAGCACTTAACTAATCG 1171
Qy      490 CCAAAAAATCTCCGAATGCAATTAACAGAAATCTTATAGTACAGGTACAGATATGCG 549
Db      1170 AAAAAAATCTTAAAGTCAATTAACAGAAATCTTATAGTACAGGTACAGATATGCG 1112
Qy      550 TCGTTAAGAGGGGCTCTAAGAAAGCACTGCTAAGTATGCAACTGTGAGATGGCC 609
Db      1111 TCAGTCAGAGGGGCTCTAAGAAAGCACTTAAATGATTAAGTATGCAACTGACAAAC 1052
Qy      610 AGTTTAAATATGACTCAACGCCCATCTGGGAGGAGCAAGAGGGGAGGGGCTCA 669
```

Db 1051 AGTTAAAGATGAATTAATGCCCAATTTGGGAGGCATGGCAGCTGTAAGAGAAAGAA 992  
QY 670 AGAGAGACTGATTAAGATGGCCATTTTGTATCTACTGTTTACAGAAATTAACCTTA 729  
Db 991 AAGCTTAAGAAATCTTCTGATTAATACCAACCTTTCTTTCAATCATCTGATTTGA 932  
QY 730 AAAAGCTTACCGGTGACACTTTTATTCAGTTGA--TTACTGCATGTCAATGTAGTG 786  
Db 931 CAGAAATTAACCTTTTAAAGTTTAACTGACACTTTCCTGCTGCAATGTAGTG 872  
QY 787 TAAATTAATCTACTACTATATA-TTAACTCAAAATACGTGTGTCTGCTTATGACGTGCT 845  
Db 871 TA-----AATCTCACTCTGATTTTGTCAAAATACGTCTTTGTCTTTATCA----- 822  
QY 846 GTTTCACACACTCCACCCAGCAGCCACGCTAGGAAGAAATACCTGCTTGAAGGCAA 905  
Db 821 ----CACACACCCACCCGACACCCACAGCTA--AACAGAAATTTCTAATGAGGAAA 768  
QY 906 CACAGAGCCAGAGTTCTGTTCAAAAGCCTGACAGACCGGTGCTGTTTGAAGAA 965  
Db 767 TAG-----CAGTTCTGTTCAAAATCTCCGCAAAAGCTGTCAGAAAACCTGCTATGAAA 714  
QY 966 CTCACATGAATCAAAAGACAGAGCTGTTACACCCATCTGACGTACAGTACAAAGTTA 1025  
Db 713 TCACAAAGACTGATTCACAAAGAGCTGAGCTGCTACGCTACATTCATACATAGTA 654  
QY 1026 CGTAATGAGCATGGGCTGATTAAGTACAGTGGCTTACATGCGCTGTGATTAAGAG 1085  
Db 653 TGTGCGGGAACACGTGCTGCTACTGCTGCTGATGATTAAGGCAACCTTCTGCGGAG 594  
QY 1086 GCTGTGCTGTGCACACGGTCTGG-----GAGCTACGGGAGGCTGTGACC--CCTGAG 1137  
Db 593 GCTGTCTGCTTTACGATCTGAGAACTACATAGGAGCAAGTGTGTGCACTCTTAC 534  
QY 1138 CCCAAGCTGCACTCTTCTTAAGGACAAAGCTCTCAACAGCTTACTGCTACGTGTC 1197  
Db 533 TCGAAGAGTACCGTCTTCTCAAAAGAGAGGCTTTCGAAAGTTCAGCTGCTGCTTC 474  
QY 1198 TCAGCACAAC---GCAACTTACTGACAGATTTTGGCAATTTCTTAATCTGAGCAGA 1254  
Db 473 TCGGCAACAAATGACAGTGTAGTTCAGAAAGTATTTTGGCAACCTTAACTGAAACA 414  
QY 1255 ATAGGGG 1261  
Db 413 ATGGGGG 407

RESULT 4  
US-10-153-668-46/c  
; Sequence 46, Application US/10153668  
; Publication No. US20030092616A1  
; GENERAL INFORMATION:  
; APPLICANT: HONDA, Goichi  
; APPLICANT: MATSUDA, Akio  
; APPLICANT: MURAMATSU, Shuji  
; APPLICANT: ISHIZAWA, Kenya  
; TITLE OF INVENTION: STAT6 Activating Gene  
; FILE REFERENCE: 1254-0207P  
; CURRENT APPLICATION NUMBER: US/10/153, 668  
; CURRENT FILING DATE: 2002-05-24  
; PRIOR APPLICATION NUMBER: US 60/293,172  
; PRIOR FILING DATE: 2001-05-25  
; PRIOR APPLICATION NUMBER: US 60/316,031  
; PRIOR FILING DATE: 2001-08-31  
; PRIOR APPLICATION NUMBER: US 60/328,403  
; PRIOR FILING DATE: 2001-10-12  
; PRIOR APPLICATION NUMBER: JP 2001-157043  
; PRIOR FILING DATE: 2001-05-25  
; PRIOR APPLICATION NUMBER: JP 2001-260681  
; PRIOR FILING DATE: 2001-08-30  
; PRIOR APPLICATION NUMBER: JP 2001-313175  
; PRIOR FILING DATE: 2001-10-10  
; NUMBER OF SEQ ID NOS: 488

; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 46  
; LENGTH: 1646  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (109)..(273)  
US-10-153-668-46  
Query Match 34.0%; Score 430.6; DB 14; Length 1646;  
Best Local Similarity 67.8%; Pred. NO. 1e-107;  
Matches 859; Conservative 0; Mismatches 324; Indels 84; Gaps 15;  
QY 50 ATGATTCAGAGATTTTAATAGTATCATGATCAAAACATCTGCTAATGATAGCAAAA 109  
Db 1646 ATGATTCAGAGATTTTAATAGTATCATGATCAAAACATCTGCTAATGATAGCAAAA 1587  
QY 110 GATCAATGTAAAAACACTCCACAAATTCGCACTGTCAATGTAAAAAGTTTGTCTAGT 169  
Db 1586 GATCAATGTAAAAACACTCCACAAATTCGCACTGTCAATGTAAAAAGTTTGTGTAGT 1527  
QY 170 GGTCAAAAGGCCCAACACTGTGTTCTTCCAGTGTAGTGTGACAGAAAGCGCTTAA 229  
Db 1526 GGTCAAAAGGTTCCACGCTGTATTTCTGCGCAGTGTAGTGTGACAGAAACATGCTCAG 1467  
QY 230 CACTAGCGC--TTGACAGAACTTCACAGACCCAAAG-----TA 266  
Db 1466 CACTAGCAGCTTTTACAGAACTTCACAGACCCAAAGCAATCAATAGCCAAAGCAGCTA 1407  
QY 267 CCGGAAGCATGTGTCGCGCGGTGAGGTGTAGAGGGGGGGGATCAATCATGACAGT 326  
Db 1406 CAGAGGCGTGTGTCGCGGTGAGGTGTAGAGGGGGGGGATCAATCATGACAGT 1348  
QY 327 GTTGTACTCTGCGCAAGACAGTATG-----TTTCAGAAATCTTAAAA 369  
Db 1347 GTGCGTATATCTGGCAAGACAGTATGTTAGAAAGTTTCATAGTTTAAGATTTCTTAAA 1288  
QY 370 TAGTTTAAAAACGTAAAGCCGACAGCTGTATTTTACACCCAGTTACTAGAAAGAA 429  
Db 1287 TATTTTAAAAACGTAAAGCCGACAGTATGTTTAACTTACTAGTAAAGAACTAA 1228  
QY 430 GGGAGCAGTATGTCAGTGTATGAAAGAGTAAAGAGTAAAGAGCACTCTCTATCTA 489  
Db 1227 GGAAGCAGCTTATATGCTCTGAATA-----AAGTACATGGAAGACCTTTACTAATCG 1173  
QY 490 CCAAAAAATCTCCGATTCATTATCAGAAAGATCTTATAGTACAGGTACAGATATGTC 549  
Db 1172 ACAAAAAACCTTCTAATGATATATCAGAAAGATTTTATATACAAAG--GAGGATATGTC 1114  
QY 550 TCGTTAAGAAAGGGGCTCTTAAAGAAAGCACTTGTCTTATGCAACTGTGAGATGCGC 609  
Db 1113 TCAGTCAGAAAGGGGCTCTATTAAGAAAGCACTTACTTAACTTACGATTAACAACAC 1054  
QY 610 AGTTTAAATATGAGACTCAACGCCCATCTGGGAGGAGACAGAGGGGGAAGGGGGCTCA 669  
Db 1053 AGTTTAAAGTATGAATTAATGCCCCAATTTGGGAGGAGTGGCAGGTGTAAAGAAAGAA 994  
QY 670 AGAGAGCACTGATTAAGATGGCCATTTGTCTACTGTGTTTCAAGAAATTAACCGTTA 729  
Db 993 AAGCTTAAGAAACATTTCTGATTAATACCAACCTTCTTCACTACTCATCTCATTTGA 934  
QY 730 AAAAGCTTACCGGTGACACTTTTATTCAGTTGA--TTACTGCATGTCAATGTAGTG 786  
Db 933 CAGAAATTAACCTTTTAAAGTTTAACTGACACTTTCCTGCTGCAATGTAGTG 874  
QY 787 TAAATTAATCTACTACTATATA-TTAACTCAAAATACGTGTGTCTGCTTATGACGTGCT 845  
Db 871 TA-----AATCTCACTCTGATTTTGTCAAAATACGTCTTTGTCTTTATCA----- 824  
QY 846 GTTTCACACACTCCACCCAGCAGCCACGCTAGGAAGAAATACCTGCTTGAAGGCAA 905  
Db 823 ----CACACACCCACCCGACACCCACAGCTA--AACAGAAATTTCTAATGAGGAAA 770





```

RESULT 8
US-09-878-178-1654
: Sequence 1654, Application US/09878178
: Patent No. US20020177552A1
: GENERAL INFORMATION:
: APPLICANT: Jiang, Yugu
: APPLICANT: Harlocker, Susan L.
: APPLICANT: Secrist, Heather
: TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
: TITLE OF INVENTION: AND DIAGNOSIS OF COLON CANCER
: FILE REFERENCE: 210121.527
: CURRENT APPLICATION NUMBER: US/09/878,178
: CURRENT FILING DATE: 2001-06-08
: NUMBER OF SEQ ID NOS: 2237
: SOFTWARE: PastSeq for Windows Version 4.0
: SEQ ID NO 1654
: LENGTH: 602
: TYPE: DNA
: ORGANISM: Homo sapien
: FEATURE:
: NAME/KEY: misc.feature
: LOCATION: (1)...(602)
: OTHER INFORMATION: n = A,T,C or G
US-09-878-178-1654

```

[illegible]

```

RESULT 9
US-10-046-935-1654
; Sequence 1654, Application US/10046935
; Publication No. US2002015601A1
; GENERAL INFORMATION:
; APPLICANT: Jiang, Yugu
; APPLICANT: Harlocker, Susan L.
; APPLICANT: Secrist, Heather
; APPLICANT: Wang, Aijun
; APPLICANT: Stolk, John A.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF COLON CANCER
; FILE REFERENCE: 210121.527C1
; CURRENT APPLICATION NUMBER: US/10/046,935
; CURRENT FILING DATE: 2002-01-15
; NUMBER OF SEQ ID NOS: 2239
; SOFTWARE: fastSeq for Windows Version 4.0
; SEQ ID NO 1654
; LENGTH: 602
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 569, 578
; OTHER INFORMATION: n = A,T,C or G
US-10-046-935-1654

```

	Query Match	15.3%	Score 197.2;	DB 13;	Length 602;
	Best Local Similarity	70.5%;	Pred. No. 1.1e-42;		
Matches	371; Conservative	0;	Mismatches 115;	Indels 40;	Gaps 7
QY	265 TACCGAAGCATGTGTCGCCGCTGGGTGAGGTCATAGAGGGGGCCGTCATCATCATGACA	324			
Db	74 TACAGGAGGCCGTGTGTCCCGCTGGGCGAGGTAAAG - GGGTCAGATTGGTCAAGTGCCA	132			
QY	325 GTGTTGGTACTCTGGCAGACAGTGTTCAGAA ------TATCTAA	367			
Db	133 GTGTGGTAATCTGGCCAGACAGTAGTATTGAAGAAGTTTCATAGTTTAAGAAATTATCTAA	192			
QY	368 AATACTTTAAAAACGTAAAGCCGCAGCACGCGATTCTACACCACGTTACTAGAAAAG	427			
Db	193 AATATTTTAAAAACATATAAAGTGCACACACATGATTTTACACCTAGTTACTAGAAAAGT	252			
QY	428 AAGGAGACACTAGTCAGCTGAGTAAAGGAAGGTGAAAACAGAACGACCTTCTACTATC	487			
Db	253 AAGGAAAGCACTTATTAGCTCTGAATA-----AAGTAACATGGAAGGACCTTTTACTAAT	307			
QY	488 TAACAAAAAAATCTCCGATGATGATTATTCAGAAAGATCTTATAGTACAGCTACATATT	547			
Db	308 CGACAAAAAAACCTCTATATGATTAATCAGAAAGATTTATPATATCAAG -GAGGCATATT	366			
QY	548 GCTGCTTAAAGAGGGGGTCTTAAAGAAAAGCACTTGCTTAAGTTTAGCACTGTGAGATGG	607			
Db	367 GCTCAGTCAAGAGGGGTTCTATTAAAGAAAGCACTTACTAAGTTAGGATGCACTTAACAAACA	426			



```

RESULT 12
US-10-066-543-2964
; Sequence 2964, Application US/10066543
; Publication No. US20030087818A1
; GENERAL INFORMATION:
; APPLICANT: Jiang, Yugu
; APPLICANT: Pyle, Ruth A.
; APPLICANT: Xu, Jiangchun
; APPLICANT: Indrias, Carol Joseph
; APPLICANT: Lodes, Michael J.
; APPLICANT: Secrist, Heather
; APPLICANT: Carter, Darick
; APPLICANT: Fanger, Gary R.
; APPLICANT: Smith, Carole L.
; APPLICANT: Durham, Margarita
; APPLICANT: Stolk, John A.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF COLON CANCER
; FILE REFERENCE: 210121.563
; CURRENT APPLICATION NUMBER: US/10/066,543
; CURRENT FILING DATE: 2002-01-31
; NUMBER OF SEQ ID NOS: 3417
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2964
; LENGTH: 567
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 527
; OTHER INFORMATION: n = A,T,C or G
US-10-066-543-2964

Query Match      14.8%; Score 187.8; DB 14; Length 567;
Best Local Similarity 73.3%; Pred. No. 3.2e-41;
Matches 304; Conservative 0; Mismatches 87; Indels 24; Gaps 4;

OY 265 TACCGAAGCATGTGTCGCCGCTGGGTGAGTCTAGAGGGGGGCGCATCATCATGAGA 324
DB 74 TACAGAGAGCGGTGTGTCGCCGCTGGGCGAGTAAGA-GGGTCAGTGTGTGTCAGTACA 132
OY 335 GTGTGTGTCCTGCGCAAGACAGTATGTTTCAGAA-----TATCTAA 367
DB 133 GTGTGCTAATCTGCGCAAGACAGTATGTTTAAAGTTTCAATAGTTTAAATATCTAA 192
OY 368 AATATTTTAAACCTGTAAAGCGGACGACGTGATTTCTACACCCAGTTACTAGAAACG 427
DB 193 AATATTTTAAACCTGTAAAGCGGACGACGTGATTTTACACCTAGTTACTAGAAACG 252
OY 428 AAGGAGACACTAGTACAGCTGATTAAGAGAGTGAAGAAACAGACGCACTTCTACTATC 487
DB 253 AAGGAGACACTATTAATAGCTCTGAATA-----AAGTAACATGGAAGCACTTTTACTAT 307
OY 488 TACCAAAAAAATCTCCGAATGCAATTATCAGAAAGATCTTATAGTACAGGTGACATAT 547
DB 308 CGACAAAAAACCCTCTAATGCAATATATCAGAAAGATTTTATATATCAAG-GAGGCATAT 366
OY 548 GTCCTTTAAGAGGGGGCTCTAAAGAAAGCACTTCTAAGTTAGCAACTGTGAGATG 607
DB 367 GTCAGTACAGAGGGGCTCTAATAGAAAGCACTTCTAAGTTAGCAACTGAGAAACA 426
OY 608 CCAGTTTAAATGAGACTCAACGCCCATCTGGGGAGGAGACAGAGGGGGAAGG 662
DB 427 CCAGTTTAAAGATGATTAATATGCCCAATTTGGGAGGACATGGCAGGTAAAG 481

```

```

; APPLICANT: ENDEGE, WILSON O., ET AL.
; TITLE OF INVENTION: NOVEL HUMAN GENES AND GENE EXPRESSION
; TITLE OF INVENTION: PRODUCTS: II
; FILE REFERENCE: CCDA-260XX
; CURRENT APPLICATION NUMBER: US/09/871,161
; CURRENT FILING DATE: 2001-05-31
; PRIOR APPLICATION NUMBER: 09/328,111
; PRIOR FILING DATE: 1999-06-08
; PRIOR APPLICATION NUMBER: 60/117,393
; PRIOR FILING DATE: 1999-01-27
; PRIOR APPLICATION NUMBER: 60/098,639
; PRIOR FILING DATE: 1998-08-31
; NUMBER OF SEQ ID NOS: 544
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 483
; LENGTH: 605
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(605)
; OTHER INFORMATION: n = A,T,C or G
US-09-871-161-483

Query Match      14.8%; Score 187.2; DB 11; Length 605;
Best Local Similarity 65.5%; Pred. No. 4.8e-41;
Matches 348; Conservative 0; Mismatches 158; Indels 25; Gaps 5;

OY 265 TACCGAAGCATGTGTCGCCGCTGGGTGAGTCTAGAGGGGGGCGCATCATCATGAGA 324
DB 74 TACAGAGAGCGGTGTGTCGCCGCTGGGCGAGTAAGA-GGGTCAGTGTGTGTCAGTACA 132
OY 325 GTGTGTGTCCTGCGCAAGACAGTATGTTTCAGAA-----TATCTAA 367
DB 133 GTGTGCTAATCTGCGCAAGACAGTATGTTTAAAGTTTCAATAGTTTAAATATCTAA 192
OY 368 AATATTTTAAACCTGTAAAGCGGACGACGTGATTTCTACACCCAGTTACTAGAAACG 427
DB 193 AATATTTTAAACCTGTAAAGCGGACGACGTGATTTTACACCTAGTTACTAGAAACG 252
OY 428 AAGGAGACACTAGTACAGCTGATTAAGAGAGTGAAGAAACAGAAAGCACTTCTACTATC 487
DB 253 AAGGAGACACTATTAATAGCTCTGAATA-----AAGTAACATGGAAGCACTTTTACTAT 307
OY 488 TACCAAAAAAATCTCCGAATGCAATTATCAGAAAGATCTTATAGTACAGGTGACATAT 547
DB 308 CGACAAAAAACCCTCTAATGCAATATATCAGAAAGATTTTATATATCAAG-GAGGCATAT 366
OY 548 GTCCTTTAAGAGGGGGCTCTAAAGAAAGCACTTCTAAGTTAGCAACTGTGAGATG 607
DB 367 GTCAGTACAGAGGGGCTCTAATAGAAAGCACTTCTAAGTTAGCAACTGAGAAACA 426
OY 608 CCAGTTTAAATGAGACTCAACGCCCATCTGGGGAGGAGACAGAGGGGGAAGG 667
DB 427 CCAGTTTAAAGATGATTAATATGCCCATTTGGGAGGAGCATGGGCTTTAAGAAAN 486
OY 668 CAAGAGACACTGATTAAGATGCGGCATTTGTC-ATCTAGTGTGACAGAAATTAACG 726
DB 487 GAAAGCACTTAAGAAAGCACTTCTGCTATATCAAAACCTTTTATTTTCTACTAGAT 546
OY 727 TTAAGAAAGCTTACCGTGACACTTTTATTAAGTTAGTGAATTAATCACTGATGAC 777
DB 547 TGACAAAAATTAACCTTTTAAAGTTTAAAGTTTAAAGTTTAAAGTTTAAAGTTTAAAG 597

```

```

RESULT 13
US-09-871-161-483
; Sequence 483, Application US/09871161
; Publication No. US20030097666A1
; GENERAL INFORMATION:

```



```
; APPLICANT: Inditias, Carol Joseph
; APPLICANT: Lodes, Michael J.
; APPLICANT: Secrist, Heather
; APPLICANT: Carter, Darick
; APPLICANT: Fanger, Gary R.
; APPLICANT: Smith, Carole L.
; APPLICANT: Durham, Margareta
; APPLICANT: Stolk, John A.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; FILE REFERENCE: 210121.563
; CURRENT APPLICATION NUMBER: US/10/066,543
; CURRENT FILING DATE: 2002-01-31
; NUMBER OF SEQ ID NOS: 3417
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3022
; LENGTH: 477
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-066-543-3022
```

```
Query Match          14.6% Score 185.4; DB 14; Length 477;
Best Local Similarity 73.2%; Pred. No. 1.3e-40;
Matches 301; Conservative 0; Mismatches 86; Indels 24; Gaps 4;

QY 265 TACCGAAGCATCTGTCGCCGCTGGGTGAGGCTTAGAGGGGGCGGCATCAATCAGATGACA 324
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 74 TACAGAGGGGCTGTGTCGCCGCTGGGTGAGGCTTAGAAG-GGGTCAATATGCTCAAGTGACA 132
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 325 GTGTGCTACTCTGGCAGACAGATGATGTTTCAGAA-----TATCTAA 367
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 133 GTGTGCTACTCTGGCAGACAGATGATGTTTCAGAA-----TATCTAA 192
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 368 AATAGTTTAAACCTGTAAGCCGACAGCATGATTTCTACCCAGTACTAGTAAACG 427
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 193 AATATTTTAAACCTGTAAGCCGACAGCATGATTTTACCTACTAGTAAACG 252
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 428 AAGGGAAGCATGCTGTCAGATGAAGGTAAGAAAGCAACCATCTTCTACTATC 487
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 253 AAGGGAAGCATGCTGTCAGATGAAGGTAAGAAAGCAACCATCTTCTACTATC 307
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 488 TACCAAAAGAAATCTCCGAATGATATATGTAAGAAAGATCTTATAGTACAGTCAACATATT 547
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 308 CGACAAAGAAACCTTCTAATGATATATGTAAGAAAGATCTTATAGTACAG- GAGGCATATT 366
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 548 GCTGCTTAAGAGGGGGCTCTAAGAAAGCACTTGTCTAAGTACCACTGTGAGATGG 607
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 367 GCTGCTTAAGAGGGGGCTCTAAGAAAGCACTTGTCTAAGTACCACTGTGAGATGG 426
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 608 CCAGTTTAAATATGACTCAAGCCCATCTGGGGAGGAGACAGCAGGGGGA 658
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 427 CCAGTTTAAATATGACTCAAGCCCATCTGGGGAGGAGACAGCAGGGGGA 477
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
```

```
RESULT 15
US-09-960-352-7612/C
; Sequence 7612, Application US/09960352
; Patent No. US20020137139A1
; GENERAL INFORMATION:
; APPLICANT: Warren, Wesley C.
; APPLICANT: Tao, Ningbing
; APPLICANT: Byatt, John C.
; APPLICANT: Mathialagan, Nagappan
; TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND
; FILE REFERENCE: 16511.006/37-21(10298)C
; CURRENT APPLICATION NUMBER: US/09/960,352
; CURRENT FILING DATE: 2001-09-24
; NUMBER OF SEQ ID NOS: 15112
; SEQ ID NO 7612
; LENGTH: 432
; TYPE: DNA
; ORGANISM: Bos taurus
```

```
; OTHER INFORMATION: Clone ID: 33-LIB188-020-01-E1-A2
US-09-960-352-7612
Query Match          13.6% Score 172.2; DB 10; Length 432;
Best Local Similarity 72.2%; Pred. No. 5.3e-37;
Matches 294; Conservative 0; Mismatches 103; Indels 10; Gaps 5;
```

```
QY 379 AACTGTAAAGCCGACGACGTGATTTCTA-CACCCAGTACTAGTAAAGCAAGGAGACA 437
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 432 AACCATTAAGCTCAGACACGTGATTTTATTAACCTAGTACTAGTAAAGCAAGGAGACA 373
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 438 CTAGTCAGC--TGAGTAAAGGAGTGAAGAAAGCAAGGACGACTTCTACTATACCAAAA 495
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 372 CTATTTAGCTTTGAATTAAGCAACATGAAGAAAGGAGTGCATTTTACTTAACCTACAAA 313
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 496 AATCTCCGAATGATATATGTAAGAAAGTCTTATAGTACAGAGTCAACATATTGCTGTTA 555
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 312 AATTTTCTAATGATATATGTAAGAAAGTCTTATATATACAAAG-GAGGCATATTGCTCATTA 254
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 556 AGAAGGGGTCTTAAGAAAGCACTTGTCTAAGTGAACACTGTGAGATGCGCAGTTTA 615
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 253 AGAATGAGTCTTAATGAAGAAAGCACTTACTAAGTGAACACTATGAGATGACAGGTCA 194
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 616 AATATGAGTCAACGCC---CCATCTGGGGAGGAGACAGCAGGGGGAAGGGGGCTCAAG 671
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 193 GAGATGATTTAAATGCCCAATTTTCAGAAAGGGGTGGCGAGTTTAAGGAAAGCTTAAGAAA 134
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 672 AGAGACACTGATTAAGATCGGCCATTTGTCTACTAGT--TTTGACAGAAATTAACGGTTA 729
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 133 ACACCTACATGATTAATACCGACCTTTCTCATCTACACGCAATTTACAGAAATTAACCTTTT 74
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 730 AAAAGCTTTACCGGTGACACTTTTATTTCACTGTAATTAATTAATTAATTAATTAATTA 776
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 73 AAAAGCTTTACCGGTGACACTTTTATTTAATTAATTAATTAATTAATTAATTAATTAATTA 27
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
```

Search completed: August 26, 2003, 02:45:51  
Job time : 315.402 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 25, 2003, 21:56:10 : Search time 3752.19 Seconds

(without alignments)  
11088.207 Million cell updates/sec

Title: US-09-717-321a-15

Perfect score: 1017  
Sequence: 1 cccattctctgctcagatc.....ccttggtctgctgaggttc 1017

Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 2888711 seqs, 2045481386 residues

Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

Database :

GenBdb1:\*

1: gb\_da:\*

2: gb\_htg:\*

3: gb\_in:\*

4: gb\_om:\*

5: gb\_ov:\*

6: gb\_pat:\*

7: gb\_ph:\*

8: gb\_pl:\*

9: gb\_pr:\*

10: gb\_ro:\*

11: gb\_sts:\*

12: gb\_sy:\*

13: gb\_un:\*

14: gb\_vl:\*

15: em\_ba:\*

16: em\_fun:\*

17: em\_hum:\*

18: em\_in:\*

19: em\_mu:\*

20: em\_om:\*

21: em\_or:\*

22: em\_ov:\*

23: em\_pat:\*

24: em\_ph:\*

25: em\_pl:\*

26: em\_ro:\*

27: em\_sts:\*

28: em\_un:\*

29: em\_vl:\*

30: em\_htg\_hum:\*

31: em\_htg\_in:\*

32: em\_htg\_other:\*

33: em\_htg\_mus:\*

34: em\_htg\_pln:\*

35: em\_htg\_rnd:\*

36: em\_htg\_mam:\*

37: em\_htg\_vrt:\*

38: em\_sy:\*

39: em\_htgo\_hum:\*

40: em\_htgo\_mus:\*

41: em\_htgo\_other:\*

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1017	100.0	1017	6 AX163751	AX163751 Sequence
2	1017	100.0	1266	2 AX163738	AX163738 Sequence
3	1013.8	99.7	217700	2 AC106124	AC106124 Rattus no
4	679.2	66.8	2281	10 BC051053	BC051053 Mus muscu
5	677.6	66.6	2319	10 BC003828	BC003828 Mus muscu
6	677.6	66.6	269081	2 AC068493	AC068493 Mus muscu
7	403.2	39.6	185245	2 AC102775	AC102775 Mus muscu
8	403.2	39.6	202342	10 AL583884	AL583884 Mouse DNA
9	337.2	33.2	179685	10 AC126556	AC126556 Mus muscu
10	290.6	28.6	1022	6 AX163752	AX163752 Sequence
11	290.6	28.6	2302	9 BC050687	BC050687 Homo sapi
12	290.6	28.6	2315	9 AK054993	AK054993 Homo sapi
13	290.6	28.6	28567	9 HSA132695	AJ132695 Homo sapi
14	290.6	28.6	212827	9 AC009412	AC009412 Homo sapi
15	285.2	28.0	1232	6 AX408985	AX408985 Sequence
16	285.2	28.0	1232	9 HDMPO25T9	D25274 Homo sapien
17	194.6	19.1	202565	9 AL354696	AL354696 Human DNA
18	193.2	19.0	605	6 AX341407	AX341407 Sequence
19	190.2	18.7	192498	2 AC105979	AC105979 Mus muscu
20	190.2	18.7	230015	2 AC132602	AC132602 Mus muscu
21	173	17.0	137625	9 AC104663	AC104663 Homo sapi
22	169.8	16.7	3544	9 AF542527	AF542527 Homo sapi
23	116.2	11.4	137625	9 AC104663	AC104663 Homo sapi
24	113.8	11.2	87616	2 AC139405	AC139405 Homo sapi
25	113	11.1	348	6 AX163753	AX163753 Sequence
26	112.4	11.1	64781	2 AC102025	AC102025 Mus muscu
27	105.8	10.4	101584	9 CNGS01D85	AL121655 BAC sequ
28	105.8	10.4	153943	9 AC012364	AC012364 Homo sapi
29	103.4	10.2	5544	9 AF542527	AF542527 Homo sapi
30	103.2	10.1	110816	9 AC002404	AC002404 Human Chr
31	98.4	9.7	170839	2 AC133467	AC133467 Mus muscu
32	85.6	8.4	495	6 AX341256	AX341256 Sequence
33	84.2	8.3	180303	9 AL672045	AL672045 Human DNA
34	84.2	8.3	201012	2 AC021189	AC021189 Homo sapi
35	76.2	7.5	228121	2 AC133022	AC133022 Rattus no
36	76.2	7.5	239768	2 AC112582	AC112582 Rattus no
37	76.2	7.5	245468	2 AC130746	AC130746 Rattus no
38	67	6.6	594	6 AX385363	AX385363 Sequence
39	62.8	6.2	194371	2 AC135453	AC135453 Rattus no
40	61.4	6.0	378	6 BD030694	BD030694 Sequence
41	56.8	5.6	198	11 G31747	G31747 sWSS2749 Er
42	54.2	5.3	7218	6 I66494	I66494 Sequence 14
43	52.4	5.2	162	6 AX397334	AX397334 Sequence
44	51.4	5.1	90354	9 AP001124	AP001124 Homo sapi
45	51.4	5.1	178089	2 AC010929	AC010929 Homo sapi

# ALIGNMENTS

RESULT 1

AX163751

LOCUS AX163751 1017 bp DNA linear PAT 22-JUN-2001

DEFINITION Sequence 15 from Patent WO0138579.

ACCESSION AX163751

VERSION AX163751.1 GI:14544857

KEYWORDS

SOURCE Rattus norvegicus (Norway rat)

ORGANISM Rattus norvegicus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.

REFERENCE 1

AUTHORS Gould-Rothberg,B.E., Dippio,V.A., Ramsehl,T.M. and Gerweil,R.W.

TITLE Method of identifying toxic agents using nsaid-induced differential

gene expression in liver  
Patent: WO 0138579-A 15 31-MAY-2001;  
Curagen Corporation (US)  
Location/Qualifiers  
1. .1017  
/organism="Rattus norvegicus"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:10116"

BASE COUNT 245 a 245 c 216 g 311 t  
ORIGIN

Query Match 100.0%; Score 1017; DB 6; Length 1017;  
Best Local Similarity 100.0%; Pred. No.1.2e-291;  
Matches 1017; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CCCCTATTCTTGGCTCAGATTAAAGATTGCCAAAATACCTGTGTGAAGTGAAGTTGGCTGTG 60  
1 CCCCTATTCTTGGCTCAGATTAAAGATTGCCAAAATACCTGTGTGAAGTGAAGTTGGCTGTG 60  
Db 1 CCCCTATTCTTGGCTCAGATTAAAGATTGCCAAAATACCTGTGTGAAGTGAAGTTGGCTGTG 60

QY 61 CTGGAACAGGTAGACACTAAGCTGTGAGAGACTTTGTCTTAAAGAGACTGCAGCTTC 120  
61 CTGGAACAGGTAGACACTAAGCTGTGAGAGACTTTGTCTTAAAGAGACTGCAGCTTC 120  
Db 61 CTGGAACAGGTAGACACTAAGCTGTGAGAGACTTTGTCTTAAAGAGACTGCAGCTTC 120

QY 121 TGGGCTCAGGGGTGCAGACCCTCCCGTAGCTCCAGACCGGTGACACAGCAGCCTCC 180  
121 TGGGCTCAGGGGTGCAGACCCTCCCGTAGCTCCAGACCGGTGACACAGCAGCCTCC 180  
Db 121 TGGGCTCAGGGGTGCAGACCCTCCCGTAGCTCCAGACCGGTGACACAGCAGCCTCC 180

QY 181 TTAATGACAGCTGCCATGTACGCCACCTGTAATCACTATACAGCCGATGCTATACGTAC 240  
181 TTAATGACAGCTGCCATGTACGCCACCTGTAATCACTATACAGCCGATGCTATACGTAC 240  
Db 181 TTAATGACAGCTGCCATGTACGCCACCTGTAATCACTATACAGCCGATGCTATACGTAC 240

QY 241 TTGTGACTGATGCTCAGAGTGGGTGAACAGCTGCTGCTTGTGATTCATATAGTAGTCT 300  
241 TTGTGACTGATGCTCAGAGTGGGTGAACAGCTGCTGCTTGTGATTCATATAGTAGTCT 300  
Db 241 TTGTGACTGATGCTCAGAGTGGGTGAACAGCTGCTGCTTGTGATTCATATAGTAGTCT 300

QY 301 CTAAATATCAGACGTGACCGGCTTCTGCAGGCTTGAACAGAACTGTGGCTCTGTGTGC 360  
301 CTAAATATCAGACGTGACCGGCTTCTGCAGGCTTGAACAGAACTGTGGCTCTGTGTGC 360  
Db 301 CTAAATATCAGACGTGACCGGCTTCTGCAGGCTTGAACAGAACTGTGGCTCTGTGTGC 360

QY 361 CTCTAACGAAGTATTCGTTCTTCTGAGTGGGTGTGCTGGGTGAGAGTGTGAAACACGA 420  
361 CTCTAACGAAGTATTCGTTCTTCTGAGTGGGTGTGCTGGGTGAGAGTGTGAAACACGA 420  
Db 361 CTCTAACGAAGTATTCGTTCTTCTGAGTGGGTGTGCTGGGTGAGAGTGTGAAACACGA 420

QY 421 CGTCATCAAGGAGACAGAGATTTTGAATATGAAGTAGAGTTAATTACACTA 480  
421 CGTCATCAAGGAGACAGAGATTTTGAATATGAAGTAGAGTTAATTACACTA 480  
Db 421 CGTCATCAAGGAGACAGAGATTTTGAATATGAAGTAGAGTTAATTACACTA 480

QY 481 CATTTGACATGAGATAATTCACTGAATAAAGTGTACAGGGTAAAGCTTTTAAACGGTT 540  
481 CATTTGACATGAGATAATTCACTGAATAAAGTGTACAGGGTAAAGCTTTTAAACGGTT 540  
Db 481 CATTTGACATGAGATAATTCACTGAATAAAGTGTACAGGGTAAAGCTTTTAAACGGTT 540

QY 541 AATTTCTGTCAACAGTAGATGACAAATGGCCGATCTTATCAGTGTCTCTTGAGCCCC 600  
541 AATTTCTGTCAACAGTAGATGACAAATGGCCGATCTTATCAGTGTCTCTTGAGCCCC 600  
Db 541 AATTTCTGTCAACAGTAGATGACAAATGGCCGATCTTATCAGTGTCTCTTGAGCCCC 600

QY 601 CCTTCCCCCTGCTGCTCCCTCCCGAGATGGGGCTGTGATTCATTTTAACTGGCCATCC 660  
601 CCTTCCCCCTGCTGCTCCCTCCCGAGATGGGGCTGTGATTCATTTTAACTGGCCATCC 660  
Db 601 CCTTCCCCCTGCTGCTCCCTCCCGAGATGGGGCTGTGATTCATTTTAACTGGCCATCC 660

QY 661 TCACAGTTGCTAACTTGAAGTGTCTTTTAAAGACCCCTCTTAAAGAGCAATATG 720  
661 TCACAGTTGCTAACTTGAAGTGTCTTTTAAAGACCCCTCTTAAAGAGCAATATG 720  
Db 661 TCACAGTTGCTAACTTGAAGTGTCTTTTAAAGACCCCTCTTAAAGAGCAATATG 720

QY 721 TCTGACCTGTACTAATAGATCTTCTGATATGCAATCGGAGATTTTGTGGTAGATAGT 780  
721 TCTGACCTGTACTAATAGATCTTCTGATATGCAATCGGAGATTTTGTGGTAGATAGT 780  
Db 721 TCTGACCTGTACTAATAGATCTTCTGATATGCAATCGGAGATTTTGTGGTAGATAGT 780

QY 781 AGAAGTGTCTTCCGTTTCACTTCTTACCTGACTAGTGTCTTCCCTGCTTTT 840  
781 AGAAGTGTCTTCCGTTTCACTTCTTACCTGACTAGTGTCTTCCCTGCTTTT 840  
Db 781 AGAAGTGTCTTCCGTTTCACTTCTTACCTGACTAGTGTCTTCCCTGCTTTT 840

QY 841 CTAGTAACCTGGGTAGAAATCAAGTGTGCGGCTTACAGTTTAACTAATTTAGAT 900  
841 CTAGTAACCTGGGTAGAAATCAAGTGTGCGGCTTACAGTTTAACTAATTTAGAT 900  
Db 841 CTAGTAACCTGGGTAGAAATCAAGTGTGCGGCTTACAGTTTAACTAATTTAGAT 900

Db 841 CTAGTAACCTGGGTAGAAATCAAGTGTGCGGCTTACAGTTTAACTAATTTAGAT 900

QY 901 ATTCTGAACATCACTGTCTTGGCCAGTAGTACCAACACTGTATGTATGATGCGGCC 960  
901 ATTCTGAACATCACTGTCTTGGCCAGTAGTACCAACACTGTATGTATGATGCGGCC 960  
Db 901 ATTCTGAACATCACTGTCTTGGCCAGTAGTACCAACACTGTATGTATGATGCGGCC 960

QY 961 CTCTAGACCTTACCCAGCGGACACATGCTTCCGGTACCTTTGGGTGTGAGGTTG 1017  
961 CTCTAGACCTTACCCAGCGGACACATGCTTCCGGTACCTTTGGGTGTGAGGTTG 1017  
Db 961 CTCTAGACCTTACCCAGCGGACACATGCTTCCGGTACCTTTGGGTGTGAGGTTG 1017

RESULT 2  
AXI63738/c  
LOCUS AXI63738 1266 bp DNA linear PAT 22-JUN-2001  
DEFINITION Sequence 2 from Patent WO0138579.  
ACCESSION AXI63738  
VERSION AXI63738.1 GI:14544844  
KEYWORDS  
SOURCE  
Rattus norvegicus (Norway rat)  
ORGANISM  
Rattus norvegicus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;  
Rattus.

REFERENCE  
1  
AUTHORS Gould-Rothberg B.E., DiIppio V.A., Ramseh, T.M. and Gervain, R.W.  
TITLE Method of identifying toxic agents using nsaid-induced differential  
gene expression in liver  
Patent: WO 0138579-A 2 31-MAY-2001;  
Curagen Corporation (US)  
LOCATION/Qualifiers  
1. .1266  
/organism="Rattus norvegicus"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:10116"

BASE COUNT 385 a 258 c 285 g 338 t  
ORIGIN

Query Match 100.0%; Score 1017; DB 6; Length 1266;  
Best Local Similarity 100.0%; Pred. No.1.3e-291;  
Matches 1017; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CCCCTATTCTTGGCTCAGATTAAAGATTGCCAAAATACCTGTGTGAAGTGAAGTTGGCTGTG 60  
1 CCCCTATTCTTGGCTCAGATTAAAGATTGCCAAAATACCTGTGTGAAGTGAAGTTGGCTGTG 60  
Db 1 CCCCTATTCTTGGCTCAGATTAAAGATTGCCAAAATACCTGTGTGAAGTGAAGTTGGCTGTG 60

QY 1201 CTGGAACAGGTAGACACTAAGCTGTGAGAGACTTTGTCTTAAAGAGACTGCAGCTTC 1142  
1201 CTGGAACAGGTAGACACTAAGCTGTGAGAGACTTTGTCTTAAAGAGACTGCAGCTTC 1142  
Db 1201 CTGGAACAGGTAGACACTAAGCTGTGAGAGACTTTGTCTTAAAGAGACTGCAGCTTC 1142

QY 121 TGGGCTCAGGGGTGCAGACCCTCCCGTAGCTCCAGACCGGTGACACAGCAGCCTCC 180  
121 TGGGCTCAGGGGTGCAGACCCTCCCGTAGCTCCAGACCGGTGACACAGCAGCCTCC 180  
Db 121 TGGGCTCAGGGGTGCAGACCCTCCCGTAGCTCCAGACCGGTGACACAGCAGCCTCC 180

QY 1141 TGGGCTCAGGGGTGCAGACCCTCCCGTAGCTCCAGACCGGTGACACAGCAGCCTCC 1082  
1141 TGGGCTCAGGGGTGCAGACCCTCCCGTAGCTCCAGACCGGTGACACAGCAGCCTCC 1082  
Db 1141 TGGGCTCAGGGGTGCAGACCCTCCCGTAGCTCCAGACCGGTGACACAGCAGCCTCC 1082

QY 181 TTAATGACAGCTGCCATGTACGCCACCTGTAATCACTATACAGCCGATGCTATACGTAC 240  
181 TTAATGACAGCTGCCATGTACGCCACCTGTAATCACTATACAGCCGATGCTATACGTAC 240  
Db 181 TTAATGACAGCTGCCATGTACGCCACCTGTAATCACTATACAGCCGATGCTATACGTAC 240

QY 1081 TTAATGACAGCTGCCATGTACGCCACCTGTAATCACTATACAGCCGATGCTATACGTAC 1022  
1081 TTAATGACAGCTGCCATGTACGCCACCTGTAATCACTATACAGCCGATGCTATACGTAC 1022  
Db 1081 TTAATGACAGCTGCCATGTACGCCACCTGTAATCACTATACAGCCGATGCTATACGTAC 1022

QY 241 TTGTGACTGATGCTCAGAGTGGGTGAACAGCTTGTCTTGAATTCATATAGTAGTCT 300  
241 TTGTGACTGATGCTCAGAGTGGGTGAACAGCTTGTCTTGAATTCATATAGTAGTCT 300  
Db 241 TTGTGACTGATGCTCAGAGTGGGTGAACAGCTTGTCTTGAATTCATATAGTAGTCT 300

QY 301 CTAAATATCAGACGTGACCGGCTTCTGCAGGCTTGAACAGAACTGTGGCTCTGTGTGC 902  
301 CTAAATATCAGACGTGACCGGCTTCTGCAGGCTTGAACAGAACTGTGGCTCTGTGTGC 902  
Db 301 CTAAATATCAGACGTGACCGGCTTCTGCAGGCTTGAACAGAACTGTGGCTCTGTGTGC 902

QY 361 CTCTAACGAAGTATTCGTTCTTCTGAGTGGGTGTGCTGGGTGAGAGTGTGAAACACGA 420  
361 CTCTAACGAAGTATTCGTTCTTCTGAGTGGGTGTGCTGGGTGAGAGTGTGAAACACGA 420  
Db 361 CTCTAACGAAGTATTCGTTCTTCTGAGTGGGTGTGCTGGGTGAGAGTGTGAAACACGA 420

QY 901 CTCTAACGAAGTATTCGTTCTTCTGAGTGGGTGTGCTGGGTGAGAGTGTGAAACACGA 842  
901 CTCTAACGAAGTATTCGTTCTTCTGAGTGGGTGTGCTGGGTGAGAGTGTGAAACACGA 842  
Db 901 CTCTAACGAAGTATTCGTTCTTCTGAGTGGGTGTGCTGGGTGAGAGTGTGAAACACGA 842

QY 421 CGTCATCAAGGAGACAGAGATTTTGAATATGAAGTAGAGTTAATTACACTA 480  
421 CGTCATCAAGGAGACAGAGATTTTGAATATGAAGTAGAGTTAATTACACTA 480  
Db 421 CGTCATCAAGGAGACAGAGATTTTGAATATGAAGTAGAGTTAATTACACTA 480

QY 841 CGTCATCAAGGAGACAGAGATTTTGAATATGAAGTAGAGTTAATTACACTA 782  
841 CGTCATCAAGGAGACAGAGATTTTGAATATGAAGTAGAGTTAATTACACTA 782  
Db 841 CGTCATCAAGGAGACAGAGATTTTGAATATGAAGTAGAGTTAATTACACTA 782

QY	481	AATGTGCATGAGTAATTCAAAGTAATTAAGAAGTCAAGGGTAAAGCTTTTAAAGGTT	540
Db	781	CATTGTACATGAGAGTAATTCACACTGTAATTAAGAAGTCAAGGGTAAAGCTTTTAAAGGTT	722
QY	541	AATTTCTGTCAACAGTAGATGACAAATAGCCGATCTTATCAGTGTCTCTTGAAGCCCC	600
Db	721	AATTTCTGTCAACAGTAGATGACAAATAGCCGATCTTATCAGTGTCTCTTGAAGCCCC	662
QY	601	CCATCCCCCGCTGTCCCTCCCGAGATGGGGCGATTGATCCATATTTAAACGGGCATCC	660
Db	661	CCATCCCCCGCTGTCCCTCCCGAGATGGGGCGATTGATCCATATTTAAACGGGCATCC	602
QY	661	TCACAGTTGCTAACCTTAGCAGATGCTTTTCTTGAAGACCCCTCTTAAAGCAGCAATATG	720
Db	601	TCACAGTTGCTAACCTTAGCAGATGCTTTTCTTGAAGACCCCTCTTAAAGCAGCAATATG	542
QY	721	TCTGACCTGTAACTTAATAGATCTTCTGTATATGATCTTGGAGATTTTTCGTAGATAGT	780
Db	541	TCTGACCTGTAACTTAATAGATCTTCTGTATATGATCTTGGAGATTTTTCGTAGATAGT	482
QY	781	AGAAGTGCCTTCCCTGTTTTCACCTTCCCTTACTAGCTAGCAATAGCTTCCCTGCTTTT	840
Db	481	AGAAGTGCCTTCCCTGTTTTCACCTTCCCTTACTAGCTAGCAATAGCTTCCCTGCTTTT	422
QY	841	CTAGTACTGTGGGTATAGAAATCACGTGCTGGGGCTTTACAGTTTAACTATTTTAGAT	900
Db	421	CTAGTACTGTGGGTATAGAAATCACGTGCTGGGGCTTTACAGTTTAACTATTTTAGAT	362
QY	901	ATTCTGAAACATCACTGCTTTCGCAGAGTACCAACACGTGATGATGATGATCGCGCCC	960
Db	361	ATTCTGAAACATCACTGCTTTCGCAGAGTACCAACACGTGATGATGATGATCGCGCCC	302
QY	961	CTCTAGACCTTACCCACGCGGACACACATGCTCCGGTACCTTTGGGTCTGTGAGGTTTC	1017
	301	CTCTAGACCTTACCCACGCGGACACACATGCTCCGGTACCTTTGGGTCTGTGAGGTTTC	245

RESULT 3				
AC106124				
LOCUS				
DEFINITION				
ACCESSION				
VERSION				
KEYWORDS				
SOURCE				
ORGANISM				
AC106124	217700 bp	DNA	linear	HTG 10-MAY-2003
Rattus norvegicus clone CH230-119E11, ***			SEQUENCING IN PROGRESS	
***, 11 unordered pieces.				
AC106124				
AC106124.6 GI:30521557				
HTG: HTGS PHASE1: HTGS_DRAFT: HTGS_ENRICHED.				
Rattus norvegicus (Norway rat)				
Rattus norvegicus				
Mammalia; Eutheria; Rodentia; Chordata; Craniata; Vertebrata; Euteleostomi;				
Rattus				
1 (bases 1 to 217700)				
Muzny,D,Marle, Metzker,M, Lee, Abramzon,S, Adams,C, Alder,J, Allen,C, Allen,H, Alsbrooks,S, Anln,A, Anguiano,D, Ayalebechil,V, Aoyagi,A, Ayodeji,M, Bacc,E, Baden,H, Baldwin,D, Bandaranaike,D, Barber,M, Barnstead,M, Bernhamed,F, Bistwal,O, Blatr,J, Blankenburg,K, Blyth,P, Brown,M, Bryant,N, Buhay,C, Burch,P, Butrell,K, Calderon,E, Cardenas,V, Carter,K, Cavazos,T, Ceasari,H, Ceron,E, Chacko,J, Chavez,D, Chen,G, Chen,R, Chen,Y, Chen,Z, Chu,J, Cleveland,C, Cockrell,R, Cox,C, Coyle,M, Cree,A, D'Souza,L, Davila,M,L, Davis,C, Davy-Carrillo,L, De Anda,C, Dedertich,D, Delgado,O, Denison,S, Deramo,C, Ding,Y, Dlinh,H, Dlyva,K, Draper,H, Dugan-Rocha,S, Dunn,A, Duplin,K, Duval,B, Eaves,K, Egan,A, Escotto,M, Eugene,C, Evans,C,A, Falls,T, Fan,G, Fernandez,S, Finley,M, Flagg,N, Forbes,L, Foster,M, Foster,P, Fraser,C,M, Gadisi,A, Ganta,R, Garcia,A, Garner,T, Garza,M, Gebregeorgis,E, Geer,K, Gill,R, Grady,M, Guerra,W, Guevara,W, Gunatareine,P, Healand,W, Hamli,C, Hamilton,K, Harvey,Y, Havlak,P, Hawes,A, Henderson,N, Hernandez,J, Hernandez,R, Hines,S, Huliy,S, Hume,J, Idlebird,D, Jackson,A, Hollins,B, Howells,S, Huliy,S, Hume,J, Idlebird,D, Jackson,A, Jackson,L, Jacob,L, Jiang,H, Johnson,B, Johnson,R, Jollivet,A, Karpathy,S, Kelly,S, Khan,Z, King,L, Kovar,C,				

TITLE	Direct Submission
JOURNAL	Unpublished
REFERENCE	2 (bases 1 to 217700)
AUTHORS	Worley, K.C.
TITLE	Direct Submission
JOURNAL	Submitted (12-JAN-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
REFERENCE	3 (bases 1 to 217700)
AUTHORS	Rat Genome Sequencing Consortium.
TITLE	Direct Submission
JOURNAL	Submitted (10-MAY-2003) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

**COMMENT**

On May 10, 2003 this sequence version replaced g1:25094619. The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (<http://www.bsc.bcm.tmc.edu/projects/rat/>). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separate by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.

Center: Baylor College of Medicine  
Center code: BCM  
Web site: <http://www.hgsc.bcm.tmc.edu/>  
Contact: [hgsc-help@bcm.tmc.edu](mailto:hgsc-help@bcm.tmc.edu)  
----- Project Information -----  
Center project name: GHL1  
Center clone name: CH230-119E11  
----- Summary Statistics -----  
Assembly program: Atlas 3.0;  
Consensus quality: 185725 bases at least Q40  
Consensus quality: 190140 bases at least Q30  
Consensus quality: 192842 bases at least Q20  
Estimated insert size: 194758; sum-of-ctrls estimation  
Quality coverage: 5x in Q20 bases; sum-of-ctrls estimation

\* NOTE: Estimated insert size may differ from sequence length  
\* (see [http://www.hgsc.bcm.tmc.edu/docs/genbank\\_draft\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html))  
\* NOTE: This is a "working draft" sequence. It currently

\* NOTE: Estimated insert size may differ from sequence length  
 \* (see [http://www.hgsc.bcm.tmc.edu/docs/Genbank\\_draft\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html))  
 \* NOTE: This is a 'working draft' sequence. It currently

\* NOTE: This is a 'working draft' sequence. It currently

\* consists of 11 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

1 109797: contig of 109797 bp in length  
\* 109798 109897: gap of unknown length  
\* 109898 166934: contig of 57037 bp in length  
\* 166935 167034: gap of unknown length  
\* 167035 176384: contig of 9350 bp in length  
\* 176385 176484: gap of unknown length  
\* 176485 206327: contig of 29843 bp in length  
\* 206328 206427: gap of unknown length  
\* 206428 207461: contig of 1034 bp in length  
\* 207462 208607: gap of unknown length  
\* 208608 208707: contig of 1046 bp in length  
\* 208708 209908: gap of unknown length  
\* 209909 210008: gap of unknown length  
\* 210009 211583: contig of 1575 bp in length  
\* 211584 211683: gap of unknown length  
\* 211684 213649: contig of 1966 bp in length  
\* 213650 213749: gap of unknown length  
\* 213750 215115: contig of 1366 bp in length  
\* 215116 215215: gap of unknown length  
\* 215216 217700: contig of 2485 bp in length.

FEATURES

1. 217700  
/organism="Rattus norvegicus"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:10116"  
/clone="CH230-119E11"

BASE COUNT 51384 a 45640 c 45954 g 51476 t 23246 others  
ORIGIN

Query Match 99.7% Score 1013.8; DB 2; Length 217700;  
Best Local Similarity 99.88; Pred. No. 1.7e-290;  
Matches 1015; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

1 CCCCTATCTTGTGCTCAGATTAAAGAAATGCCAAATACCTGTGTGAAGTTGGCTGTG 60  
63447 CCCCTATCTTGTGCTCAGATTAAAGAAATGCCAAATACCTGTGTGAAGTTGGCTGTG 63506  
QY 61 CTGGAACACGTAAAGCACTAAAGCTGTGAGAGACTTGTCTTAAAGAACTGCACTTC 120  
63507 CTGGAACACGTAAAGCACTAAAGCTGTGAGAGACTTGTCTTAAAGAACTGCACTTC 63566  
Db 121 TGGGCTCAGGGGGTGACACCCCTCCGTAAGCTCCAGACCGGTGACACAGACGCTCC 180  
63567 TGGGCTCAGGGGGTGACACCCCTCCGTAAGCTCCAGACCGGTGACACAGACGCTCC 63626  
Db 181 TTAATGACAGCGTCCATGTAAACGACCTGTAACTTAACAGCCATGCTCTTAAGTAAC 240  
63627 TTAATGACAGCGTCCATGTAAACGACCTGTAACTTAACAGCCATGCTCTTAAGTAAC 63686  
QY 241 TTGTGCTGACGTACAGATGGGTGAACAGCTGTCTTGTGATTTCAATAGAGTTCT 300  
63687 TTGTGCTGACGTACAGATGGGTGAACAGCTGTCTTGTGATTTCAATAGAGTTCT 63746  
Db 301 CTAAATACACAGCTGACCGGCTTCTGACAGGTTTGAACAGACCTGCTCTCTGTTC 360  
63747 CTAAATACACAGCTGACCGGCTTCTGACAGGTTTGAACAGACCTGCTCTCTGTTC 63806  
QY 361 CTCTAAGCAAGTATTTCTTCTAGTGTGGGTGTGGGTGGGTGGGTGGGTGGGTGGGT 420  
63807 CTCTAAGCAAGTATTTCTTCTAGTGTGGGTGTGGGTGGGTGGGTGGGTGGGTGGGT 63866  
Db 421 CGTCATCAAGAGACAGACAGATTTTGTGACTAATATGAACTAGAGATTAATTACACTA 480  
63867 CGTCATCAAGAGAGACAGACAGATTTTGTGACTAATATGAACTAGAGATTAATTACACTA 63926

QY 481 CATTGTACATGAGTAATTAATCACTGAATTAAGTGCACGGGTAAACCTTTTAACGGTT 540  
Db 63927 CATTGTACATGAGTAATTAATCACTGAATTAAGTGCACGGGTAAACCTTTTAACGGTT 63986  
QY 541 AATTCTGTAACAGAGATGACAAATGGCCGCTTATACAGTCTCTTGAGGCC 600  
63987 AATTCTGTAACAGAGATGACAAATGGCCGCTTATACAGTCTCTTGAGGCC 64046  
QY 601 CCTTCCCGCTGCTGCTCCCTCCCGATGGGGCGGTGAGTCCATTTTAACTGGCCATCC 660  
64047 CCTTCCCGCTGCTGCTCCCTCCCGATGGGGCGGTGAGTCCATTTTAACTGGCCATCC 64106  
Db 661 TCACAGTTGCTACTTACGAAGTCTTTTCTTTTAAAGACCCCTTTTAACGACCAATATG 720  
64107 TCACAGTTGCTACTTACGAAGTCTTTTCTTTTAAAGACCCCTTTTAACGACCAATATG 64166  
QY 721 TCTGACGTGACTTAAGATCTTCTGATATGATCGGAGATTTTGGTGGATAGT 780  
64167 TCTGACGTGACTTAAGATCTTCTGATATGATCGGAGATTTTGGTGGATAGT 64226  
QY 781 AGAAGTCGCTCTGCTTTCACCTTCTTACCTGCTGACTAGTGTCCCTGCTTT 840  
64227 AGAAGTCGCTCTGCTTTCACCTTCTTACCTGCTGACTAGTGTCCCTGCTTT 64286  
Db 841 CTACTACTGGGTGTAGAAATACAGTCTGCGGCTTTTACAGTTTAACTATTATGAT 900  
64287 CTACTACTGGGTGTAGAAATACAGTCTGCGGCTTTTAACTATTATGAT 64346  
QY 901 ATTCTGAACATCACTGCTCTTCCAGAGTACCAACATGTCATGTGATGTCGGGCC 960  
64347 ATTCTGAACATCACTGCTCTTCCAGAGTACCAACATGTCATGTGATGTCGGGCC 64406  
QY 961 CTCTAGACCTTACCCACGCGGACACATGCTTCCGTAACCTTTGGGTGTGAGGTT 1017  
64407 CTCTAGACCTTACCCACGCGGACACATGCTTCCGTAACCTTTGGGTGTGAGGTT 64463

RESULT 4

BC051053  
LOCUS 2281 bp mRNA linear ROD 14-APR-2003  
DEFINITION Mus musculus, clone MGC:58966 IMAGE:5038182, mRNA, complete cds.  
ACCESSION BC051053  
VERSION BC051053.1 GI:29835221  
KEYWORDS MGC.

SOURCE Mus musculus (house mouse)  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 2281)  
AUTHORS Strausberg, R.  
TITLE Direct Submission  
JOURNAL  
REMARK NIH-MGC Project URL: <http://mgc.ncl.nih.gov>  
COMMENT Contact: [cgabbs-remail.nih.gov](mailto:cgabbs-remail.nih.gov)  
Email: [cgabbs-remail.nih.gov](mailto:cgabbs-remail.nih.gov)  
Tissue Procurement: Gilbert Smith, Ph.D.  
cDNA Library Preparation: Life Technologies, Inc.  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (ILIN)  
DNA Sequencing by: National Institutes of Health Intramural  
Sequencing Center (NISC),  
Gaithersburg, Maryland;  
Web site: <http://www.nisc.nih.gov/>  
Contact: [nisc.mgc@nih.gov](mailto:nisc.mgc@nih.gov)

Akhter, N., Ayle, K., Beckstrom-Sternberg, S.M., Benjamin, B.,  
Blakesley, R.W., Bouffard, G.G., Breen, K., Brinkley, C., Brooks, S.,  
Diehl, N.L., Grantham, S., Guan, X., Gupta, J., Haghighi, P.,  
Hansen, N., Ho, S.-L., Karins, E., Kwong, P., Lalic, P., Legaspi, R.,  
Maduro, O.L., Masello, C., Masker, B., Mastrian, S.D., McCloskey, J.C.,  
McGowan, J., Pearson, R., Stantrop, S., Thomas, P.J., Touchman, J.W.,  
Tsurganov, C., Vogt, J.L., Walker, M.A., Wetherby, K.D., Wiggins, L.,

Young, A., Zhang, L.-H. and Green, E.D.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/BLIN at: <http://limage.llnl.gov>  
Series: IRAC Plate: 108 Row: f Column: 9  
This clone was selected for full length sequencing because it passed the following selection criteria: Hexamer frequency ORF analysis, Genomescan gene prediction, Similarity but not identity to protein.

#### FEATURES

source

Location/Qualifiers  
1..2281  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/strain="Czech 11"  
/db\_xref="taxon:10090"  
/clone="MGC:58966 IMAGE:5038182"  
/issue\_type="Mammary tumor metastasized to lung. Tumor arose spontaneously from a senescent normal mammary (clonal) outgrowth infected with the virus MMTV."  
/clone\_1lb="NCI\_CGAP\_1429"  
/lab\_host="DH10B"  
/note="Vector: pCMV-SPORT6"  
197..775  
/codon\_start=1  
/product="Unknown (protein for MGC:58966)"  
/protein\_id="AAH51053.1"  
/db\_xref="GI:29835222"  
/translation="MOAICVYVGDGAVGKTCILLISYTNAPPEXIPVDFNYSANV  
MYDKRVNLGLMDTQEDYDLRLPISOTDVELICFSLSPASBPASRAKRYEVR  
HRCPTPIIVSTKIDLRDQKTIKAKKILPTTYPQGLAMAKETGAKVILCESAL  
TORGKTYVDEIRAVLCPPVKKRRKKLLL"

CDS  
BASE COUNT 568 a 569 c 504 g 640 t  
ORIGIN

Query Match 66.8%; Score 679.2; DB 10; Length 2281;  
Best Local Similarity 86.1%; Pred. No. 6,7e-191;

Matches 876; Conservative 0; Mismatches 113; Indels 29; Gaps 10;

1 CCCCTATTCCTGCTCAGATTAAAGTAATGCCAAATACCTGTGTAAGTTCGTTG 60  
1019 CCCCATCTCTGCTCAGATTAAATGCAAAATACCTGTGTAAGTTCGTTG 1078  
61 CTGAGAACAGTAACTAAGCTGTGAGACA--CTTGTCTCTTAAGACAGCTCAGC 117  
1079 CTGAGAACAGTAACTAAGCTGTGAGACA--CTTGTCTCTTAAGACAGCTCAGC 1138  
118 TTCTGGGCTCAGGGGTGACAGCCCTCCGTAAGTCCAGACGCTGTGACAGCAGCC 177  
1139 TTCTGGGCTCAGGGGTGACAGCCCTCCGTAAGTCCAGACGCTGTGACAGCAGCC 1198  
178 TCCTTAATGACAGCGCTGCATGTAAAGCACTGTAACTATATGAGCCATCTTACGT 237  
1199 TCCTTAATGACAGCGCTGCATGTAAAGCACTGTAACTATATGAGCCATCTTACGT 1258  
238 AACTTTGTACTAGTCTAC---GATGGGTGTAAAGCTCTCTTATTTCAATAGTG 294  
1259 AACTTTGTACTAGTCTAC---GATGGGTGTAAAGCTCTCTTATTTCAATAGTG 1318  
295 AACTTTGTACTAGTCTAC---GATGGGTGTAAAGCTCTCTTATTTCAATAGTG 354  
1319 AACTTTGTACTAGTCTAC---GATGGGTGTAAAGCTCTCTTATTTCAATAGTG 1378  
355 TGTGCTCTTAAAGAGAGTATCTGTCTAGTCTGTGGGTGTGCTGGGTGAGTGTGAA 414  
1379 TGTGCTCTTAAAGAGAGTATCTGTCTAGTCTGTGGGTGTGCTGGGTGAGTGTGAA 1436  
415 ACACGACGTCATCAAAAGAGAGACAGACAGTATTTGACTATATGAAAGTAAATATT 474  
1437 ACACGACGTCATCAAAAGAGAGACAGACAGTATTTGACTATATGAAAGTAAATATT 1496  
475 ACACGACGTCATCAAAAGAGAGACAGACAGTATTTGACTATATGAAAGTAAATATT 534  
1497 ACACGACGTCATCAAAAGAGAGACAGACAGTATTTGACTATATGAAAGTAAATATT 1556

535 ACGTTAAATTTCTGTCAAACAGATAGACAAATAGGCGCATTTATCAGTCTC----- 589  
1557 ACGTTAAATTTCTGTCAAACAGATAGACAAATAGGCGCATTTATCAGTCTCTTGAG 1616  
590 -----TCTGAGCCGCCCTCCCTCCCTGCTGTCCTCCAGATGGGGGTGAGTCA 642  
1617 CACCCCTCCGCCGCCGCCCTCCCTCCCTGCTGTCCTCCAGATGGGGGTGAGTCA 1676  
643 TATTTAAAGTGGCAGCCAGTCCAGTCTTAACTTAACTTAACTTAACTTAACTTAA 702  
1677 TATTTAAAGTGGCAGCCAGTCTTAACTTAACTTAACTTAACTTAACTTAACTTAA 1734  
703 TTTCTTAACGAGCAATATGCTGACCTGATATATAGTCTTCTGATATGATTCG-CA 761  
1735 TTTCTTAACGAGCAATATGCTGACCTGATATATAGTCTTCTGATATGATTCG-CA 1794  
762 GATTTTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 820  
1795 TTTTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1854  
821 CTAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 880  
1855 GTTAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1914  
881 GTTTTAACTATTTTAACTATTTTAACTATTTTAACTATTTTAACTATTTTAACTAT 938  
1915 GTTTTAACTATTTTAACTATTTTAACTATTTTAACTATTTTAACTATTTTAACTAT 1974  
939 GTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 996  
1975 GTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2029

RESULT 5  
BC003828  
LOCUS  
DEFINITION  
MUS MUSCULUS RAS-RELATED C3 BOTULINUM SUBSTRATE 1, mRNA (CDNA CLONE  
ACCESSION  
BC003828  
VERSION  
BC003828.1 GI:13277917  
KEYWORDS  
MUS MUSCULUS (HOUSE MOUSE)  
SOURCE  
MUS MUSCULUS  
ORGANISM  
REFERENCE  
AUTHORS  
1 (bases 1 to 2319)  
Strausberg, R.L., Feilgold, E.A., Grouse, L.H., Derge, J.G.,  
Klausner, R.D., Collins, F.S., Wagner, L., Shennan, C.M., Schuler, G.D.,  
Altschul, S.F., Zeeberg, B., Buetow, K.H., Schaefer, C.F., Bhat, N.K.,  
Hopkins, R.F., Jordan, H., Moore, T., Max, S.I., Wang, J., Hsieh, F.,  
Datchenko, L., Marusina, K., Farmer, A.A., Rubin, G.M., Hong, L.,  
Stapleton, M., Soares, M.B., Bonaldo, M.F., Casavant, T.L.,  
Scheetz, T.E., Brownstein, M.J., Usdin, T.B., Toshiyuki, S.,  
Carninci, P., Prange, C., Raha, S.S., Loquellano, N.A., Peters, G.J.,  
Abramson, R.D., Mullany, S.J., Bosak, S.A., McEwan, P.J.,  
McKernan, K.J., Malek, J.A., Gunaratne, P.H., Richards, S.,  
Worley, K.C., Hale, S., Garcia, A.M., Gay, L.J., Huljk, S.W.,  
Villalón, D.K., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A.,  
Rahney, J., Helton, E., Kettelman, M., Madan, A., Rodriques, S.,  
Sanchez, A., Whiting, M., Madan, A., Young, A.C., Shevchenko, Y.,  
Bouffard, G.G., Blakesley, R.W., Touchman, J.W., Green, E.D.,  
Dickson, M.C., Rodriguez, A.C., Grimwood, J., Schmitz, J., Myers, R.M.,  
Butterfield, Y.S., Kravitski, M.I., Skalske, U., Smallus, D.E.,  
Scherer, A., Schein, J.E., Jones, S.J. and Marra, M.A.  
Generation and initial analysis of more than 15,000 full-length  
human and mouse cDNA sequences  
Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)  
JOURNAL  
MEDLINE  
PUBMED  
12477932  
2 (bases 1 to 2319)  
Strausberg, R.  
REFERENCE  
TITLE  
Direct Submission





## JOURNAL

## COMMENT

Submitted (03-MAY-2000) Department of Molecular Genetics, Albert Einstein College of Medicine Genome Center, 1300 Morris Park Ave., Bronx, NY 10461, USA  
On Aug 11, 2001 this sequence version replaced gi:14993654.

Center: Harvard Partners Genome Center

Center Code: HPGC

Web site: <http://www.hpcgg.org/Sequence/mouse.html>  
Contact: [hpcgcmendel.mgh.harvard.edu](mailto:hpcgcmendel.mgh.harvard.edu)

-----Summary Statistics

Center project name: ABN  
Sequencing vector: pUC18, L08752  
Chemistry: Dye-terminator Big Dye, 100%  
\*Consensus quality: 256302 at least Q20  
\*Consensus quality: 254124 at least Q30  
\*Consensus quality: 250276 at least Q40  
Estimated insert size: agarose-FP - N/A  
\*Estimated insert size: 268581 - sum-of-contigs  
Quality coverage: agarose-FP - N/A  
Quality coverage: 6.2 x in Q20 bases; sum-of-contigs estimation

\* NOTE: This is a 'working draft' sequence. It currently consists of 26 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved

1 51730: contig of 51730 bp in length  
\* 51731 51750: gap of unknown length  
\* 51751 82423: contig of 30673 bp in length  
\* 82424 82444 118158: contig of 35715 bp in length  
\* 118159 118178: gap of unknown length  
\* 118179 139441: contig of 21263 bp in length  
\* 139442 139461: gap of unknown length  
\* 139462 160638: contig of 21177 bp in length  
\* 160639 160658: gap of unknown length  
\* 160659 177611: contig of 16953 bp in length  
\* 177612 177632: gap of unknown length  
\* 177633 190991: contig of 13360 bp in length  
\* 190992 191011: gap of unknown length  
\* 191012 200128: contig of 9117 bp in length  
\* 200129 200148: gap of unknown length  
\* 200149 211882: contig of 11734 bp in length  
\* 211883 211902: gap of unknown length  
\* 211903 220773: contig of 8871 bp in length  
\* 220774 220793: gap of unknown length  
\* 220794 229663: contig of 8850 bp in length  
\* 229664 229663: gap of unknown length  
\* 229664 238339: contig of 8676 bp in length  
\* 238340 238359: gap of unknown length  
\* 238360 244848: contig of 6489 bp in length  
\* 244849 244868: gap of unknown length  
\* 244869 249725: contig of 4857 bp in length  
\* 249726 249745: gap of unknown length  
\* 249746 252523: contig of 2778 bp in length  
\* 252524 252543: gap of unknown length  
\* 252544 257520: contig of 4977 bp in length  
\* 257521 257540: gap of unknown length  
\* 257541 260396: contig of 2856 bp in length  
\* 260397 260416: gap of unknown length  
\* 260417 262709: contig of 2293 bp in length  
\* 262710 262729: gap of unknown length  
\* 262730 263934: contig of 1205 bp in length  
\* 263935 263954: gap of unknown length  
\* 263955 264196: contig of 242 bp in length  
\* 264197 264216: gap of unknown length  
\* 264217 264629: contig of 413 bp in length  
\* 264630 265608: gap of unknown length  
\* 265609 265628: gap of unknown length

\* 265629 266935: contig of 1307 bp in length  
\* 266936 266955: gap of unknown length  
\* 266956 268335: contig of 1380 bp in length  
\* 268336 268355: gap of unknown length  
\* 268356 268549: contig of 194 bp in length  
\* 268550 268569: gap of unknown length  
\* 268570 269081: contig of 512 bp in length.  
location/Qualifiers  
1. 269081

## FEATURES

source

misc\_feature  
/note="assembly\_name:Contig138"  
51751..82423  
/note="assembly\_name:Contig137"  
82444..118158  
/note="assembly\_name:Contig136"  
118179..139441  
/note="assembly\_name:Contig135"  
139462..160638  
/note="assembly\_name:Contig134"  
160659..177611  
/note="assembly\_name:Contig133"  
177632..190991  
/note="assembly\_name:Contig132"  
191012..200128  
/note="assembly\_name:Contig131"  
200149..211882  
/note="assembly\_name:Contig130"  
211903..220773  
/note="assembly\_name:Contig129"  
220794..229643  
/note="assembly\_name:Contig128"  
229664..238339  
/note="assembly\_name:Contig127"  
238360..244848  
/note="assembly\_name:Contig126"  
244869..249725  
/note="assembly\_name:Contig125"  
249746..252523  
/note="assembly\_name:Contig124"  
clone\_end:SP6  
vector\_side:left"  
252544..257520  
/note="assembly\_name:Contig123"  
257541..260396  
/note="assembly\_name:Contig122"  
260417..262709  
/note="assembly\_name:Contig121"  
262730..263934  
/note="assembly\_name:Contig120"  
263955..264196  
/note="assembly\_name:Contig119"  
264217..264629  
/note="assembly\_name:Contig118"  
264650..265608  
/note="assembly\_name:Contig117"  
265629..266935  
/note="assembly\_name:Contig116"  
266956..268335  
/note="assembly\_name:Contig115"  
268356..268549  
/note="assembly\_name:Contig114"  
268570..269081  
/note="assembly\_name:Contig113"  
BASF COUNT 69726 a 66027 c 64672 g 67724 t 932 others  
ORIGIN  
Query Match 66.6%; Score 677.6; DB 2; length 269081;

Best Local Similarity 85.3%; Pred. No. 3e-190;  
Matches 870; Conservative 0; Mismatches 119; Indels 31; Gaps 9;

```

QY 1 CCCCTATCTTGGTCAGATTAGAAATGGCCAAATACCTGTGTAAGTTGGCTGTG 60
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 62207 CCCCCTATCTTGGTCAGATTAGAAATGGCCAAATACCTGTGTAAGTTGGCTGTG 62266
QY 61 CTGGAACACGTAAAGCACTAAGCTGTGGAGAGA--CTTGTCTTTAAGAAAGCTGAC 117
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 62267 CTGGAACACGTAAAGCACTAAGCTGTGGAGAGACTTCTGTCTTGAAGAGACGGAGC 62326
QY 118 TTTGGGGCTCAGGGGTGACACCTCCCGTATGCCAGACCGGTGACAGACAGCAGCC 177
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 62327 TTTGGAGCCAGGGGTGACACCTCCCGTATGCCAGACCGGTGACAGACAGCAGCC 62386
QY 178 TCCCTAATGACACCTGTCATGTAACGACACCTGTAACTTATCAGCCCATGCTATTACGT 237
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 62387 TCCCTAATGACACCTGTCATGTAACGACACCTGTAACTTATCAGCCCATGCTATTACGT 62446
QY 238 AACCTTTTACTGTACGTCAC--GATGGGTGTAACAGCTGCTGCTTTGATTGATAGTG 294
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 62447 AACCTTTTACTGTACGTCACGATGATAGTGTAAGCTGACGCTCTTGTGATTGATAGTG 62506
QY 295 AGTCTCTAATAAATACGACGTGACCGGCTTCTGACAGGCTTTGACAGAACTGCGCTCGT 354
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 62507 AGTCTCTAATAAATACGACGTGACCGGCTTCTGACAGGCTTTGACAGAACTGCGCTCGT 62566
QY 355 TGTTCCTCTAAGCAAGTATTCTGTCTCTAGTCTGTGGGTGTGCTGGGTGAGTGTGAA 414
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 62567 TGTTCCTCTAAGCAAGTATTCTGTCTCTAGTCTGTGGGTGTGCTGGGTGAGTGTGAA 62624
QY 415 ACAGAGCTCATCAAAAGAGACAGACGATTTTGTGACATAATATGAGAGAGATTAATTT 474
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 62625 ACAGAGCTCATCAAAAGAGACAGACGATTTTGTGACATAATATGAGAGAGATTAATTT 62684
QY 475 ACACCTAATGTTACATGGAGTAATTCACACTGAATATAAGTGTACAGGGTAAGCTTTT 534
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 62685 ACACCTAATGTTACATGGAGTAATTCACACTGAATATAAGTGTACAGGGTAAGCTTTT 62744
QY 535 ACGGTATATTTCTGTCAAAAGTAGATGACAAATGCGCATCTTATCAGTGTCTC---- 589
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 62745 ACGGTATATTTCTGTCAAAAGTAGATGACAAATGCGCATCTTATCAGTGTCTCTGAG 62804
QY 590 -----TCTTGAAGCCCCCTTCCCGCTGCTGCCCTCCCGAGNNGGGCGCTTGAAGTCA 642
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 62805 CATCCCTTCGCGCCCCCACCCCTCCACTGCGCTCCGCCAGATGGGGCGTTGAGTCCA 62864
QY 643 TATTTAAACTGGCCATCTCAGCTGTGCTAATAGCAAGTGTCTTTCTTTAGGACCCCC 702
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 62865 TATTTAAACTGGCCATCTCAGCTGTGCTAATAGCAAGTGTCTTTCTTTAGGACCCCC 62922
QY 703 TTTCTTAACGAGCAATATGTCTGACCTGTACTATTAAGATCTTTCTGATNATGCA----TT 757
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 62923 TTTCTTAACGAGCAATATGTCTGACCTGTACTATTAAGATCTTTCTGATNATGCA----TT 757
QY 758 CGGAGATTTTGTGATAGTATGTAAGTGGTGGTCTTGTACCTTCCCTTACTGAG- 816
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 62883 TTTGTTTTTTTTTGTGAGCTAGTAAAGTGGCTGCTGTTTTGGCTTACTTACTTCAAGA 63042
QY 817 CTGACTAGTGGTCCCTTCTGTTTCTAATACTGGGTGTAGAAATCAAGTCTCGGGCTT 876
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 63043 GCTGATGATGCTTCCCTTGTAGTTTCTGTAAGCACTAGTGTGCAAAATCATGTTTGCAAGTT 63102
QY 877 TACAGTTTTTAACTATTTTATGATATCTGAAACATCACTGTCTTGGCAGATGACACA 936
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 63103 TACAGTTTTTAACTATTTTATGATATCTTAACTATGAAACCTTCTAAACGCTACACA 63162
QY 937 CTGTCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 996
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 63163 CTGTCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 63219

```

RESULT 7

AC102775  
LOCUS  
DEFINITION  
MUS MUSCULUS CLONE RP23-115C10, WORKING DRAFT SEQUENCE, 9 unordered pieces.  
AC102775  
ACCESSION  
AC102775.3 GI:29164654  
VERSION  
HTG5, PHASE1; HTG5\_DRAFT; HTG5\_FULLTOP.  
KEYWORDS  
HTG; MUSCULUS (house mouse)  
SOURCE  
ORGANISM  
Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sclerogasthi; Muridae; Murinae; Mus.  
REFERENCE  
1 (bases 1 to 185245)  
Birren, B., Nussbaum, C. and Lander, E.  
Mus musculus, clone RP23-115C10  
Unpublished  
2 (bases 1 to 185245)  
Birren, B., Linton, L., Nussbaum, C., Lander, E., Ali, A., Allen, N., Anderson, S., Barna, N., Bastien, V., Boguslavsky, L., Boukhalter, B., Brown, A., Camarata, J., Campolano, A., Chang, J., Chazaro, B., Choebel, Y., Colangelo, M., Collins, S., Collymore, A., Cook, A., Cooke, P., DeArrellano, K., Dewar, K., Diaz, J. S., Dodge, S., Faro, S., Ferreira, P., FitzHugh, W., Gage, D., Galagan, J., Gardyna, S., Glade, S., Gord, S., Goyette, M., Graham, L., Grand-Pierre, N., Hagos, B., Heaford, A., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Kamat, A., Karatas, A., Kells, C., Lacroque, K., Llanazares, R., Landers, F., Lehoczy, J., Levine, R., Liu, G., Maclean, C., Macdonald, P., Major, J., Marquis, N., Matthews, C., McCarthy, M., McEwan, P., McKernan, K., McPheeters, R., Meldrim, J., Menues, L., Mihoval, T., Mlenka, V., Murphy, T., Naylor, J., Nguyen, C., Norbu, C., Norman, C. H., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Pollard, V., Raymond, C., Retta, R., Rieback, M., Riley, R., Rise, C., Rogov, P., Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, S., Schupack, R., Seaman, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N., Strauss, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J., Topham, K., Travers, M., Travis, N., Trigilio, J., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W. J., Young, C., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.  
Direct Submission  
Submitted (23-NOV-2001) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA  
3 (bases 1 to 185245)  
Birren, B., Nussbaum, C., Lander, E., Abouelkell, A., Allen, N., Anderson, S., Arachchi, H. M., Barna, N., Bastien, V., Bloom, T., Boguslavsky, L., Boukhalter, B., Camarata, J., Chang, J., Choebel, Y., Collymore, A., Cook, A., Cooke, P., Corum, B., DeArrellano, K., Diaz, J. S., Dodge, S., Dooley, K., Dorris, L., Erickson, J., Faro, S., Ferreira, P., FitzGerald, M., Gage, D., Galagan, J., Gardyna, S., Graham, L., Grand-Pierre, N., Haef, N., Haggopian, D., Hagos, B., Hall, J., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Kamat, A., Karatas, A., Kells, C., Landers, T., Levine, R., Lindblad-Toh, K., Liu, G., Lui, A., Mabbitt, R., Maclean, C., Macdonald, P., Major, J., Manning, J., Matthews, C., McCarthy, M., Meldrim, J., Menues, L., Mihoval, T., Mlenka, V., Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Norbu, C., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Rachupka, A., Ramasamy, U., Raymond, C., Retta, R., Rise, C., Rogov, P., Roman, J., Schauer, S., Schupack, R., Seaman, S., Severy, P., Smith, C., Spencer, B., Stange-Thomann, N., Stojanovic, N., Stubbs, M., Talamas, J., Tesfaye, S., Theodore, J., Topham, K., Travers, M., Vassiliev, H., Venkataraman, V. S., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Young, C., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.  
Direct Submission  
Submitted (23-MAR-2003) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA  
On Mar 23, 2003 this sequence version replaced gi:22381790.  
All repeats were identified using RepeatMasker:  
Smit, A. F. A. & Green, P. (1996-1997)  
http://ftp.genome.washington.edu/RM/RepeatMasker.html  
----- Genome Center  
Center: Whitehead Institute/ MIT Center for Genome Research  
Center code: MIBR  
Web site: http://www-seq.wi.mit.edu

Contact: sequence\_submissions@genome.wi.mit.edu  
 ----- Project Information  
 Center Project name: l19647  
 Center Clone name: l15.C.10  
 ----- Summary Statistics  
 Sequencing vector: Plasmid; n/a; 100% of reads  
 Chemistry: Dye-terminator Big Dye; 100% of reads  
 Assembly program: Phrap; version 0.960731  
 Consensus quality: 183758 bases at least Q40  
 Consensus quality: 184136 bases at least Q30  
 Consensus quality: 184291 bases at least Q20  
 Insert size: 178000; agarose-fp  
 Insert size: 184445; sum-of-fragments  
 Quality coverage: 12.6 in Q20 bases; agarose-fp  
 Quality coverage: 12.2 in Q20 bases; sum-of-fragments  
 -----  
 \* NOTE: This is a 'working draft' sequence. It currently  
 \* consists of 9 contigs. The true order of the pieces  
 \* is not known and their order in this sequence record is  
 \* arbitrary. Gaps between the contigs are represented as  
 \* runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence  
 \* as soon as it is available and the accession number will  
 \* be preserved.  
 \*  
 \* 1 53693: contig of 53693 bp in length  
 \* 53694 53793: gap of 100 bp  
 \* 53794 56675: contig of 2882 bp in length  
 \* 56676 56775: gap of 100 bp  
 \* 56776 60589: contig of 3814 bp in length  
 \* 60590 60690: gap of 100 bp  
 \* 60690 67210: contig of 6521 bp in length  
 \* 67211 67310: gap of 100 bp  
 \* 67311 82585: contig of 15275 bp in length  
 \* 82586 103118: gap of 100 bp  
 \* 103119 103218: contig of 20433 bp in length  
 \* 103219 132145: contig of 28927 bp in length  
 \* 132146 132246: gap of 100 bp  
 \* 132246 182335: contig of 50090 bp in length  
 \* 182336 182435: gap of 100 bp  
 \* 182436 185245: contig of 2810 bp in length.  
 \*  
 \* Location/Qualifiers  
 \* 1.185245  
 \* /organism="Mus musculus"  
 \* /mol\_type="genomic DNA"  
 \* /db\_xref="taxon:10090"  
 \* /clone\_1fb="RP23-115C10"  
 \* /clone\_1fb="RP23-23 Female Mouse BAC"  
 \* 1.53693  
 \* /note="assembly-fragment  
 \* clone\_end:SP6  
 \* vector\_side:left"  
 \* 53794.56675  
 \* /note="assembly-fragment"  
 \* 56776.56775  
 \* /note="assembly-fragment"  
 \* 60690.67210  
 \* /note="assembly-fragment"  
 \* 67311.82585  
 \* /note="assembly-fragment"  
 \* 82586.103118  
 \* /note="assembly-fragment"  
 \* 103219.132145  
 \* /note="assembly-fragment"  
 \* 132246.182335  
 \* /note="assembly-fragment"  
 \* 182436.185245  
 \* /note="assembly-fragment  
 \* clone\_end:T7  
 \* vector\_side:right"  
 \*  
 \* BASE COUNT 56670 a 34956 c 35069 g 57550 t 800 others  
 \* ORIGIN

Query Match 39.6%; Score 403.2; DB 2; Length 185245;  
 Best Local Similarity 89.7%; Pred. No. 1.9e-108;  
 Matches 481; Conservatve 0; Mismatches 43; Indels 12; Gaps 4;  
 QY 1 CCCCTATTCCTGCTCAGATTAGAAATGCCAAATACCTTGCACTAAGTCGTTGG 60  
 Db |||||||  
 QY 74 CCCCTATTCCTGCTCAGATTAGAAATGCCAAATACCTTGCACTAAGTCGTTGG 133  
 Db |||||||  
 QY 61 CTGAGAACGCTAAGACCTAAGCTGTTGAGAG---CTTGGCTTAAAGACCTGAC 117  
 Db |||||||  
 QY 134 CTGAGAACGCTAAGACCTAAGCTGTTGAGAGCTTGGCTTAAAGACCTGAC 193  
 Db |||||||  
 QY 118 TTCTGGCTCAGGGGTGACAGACCTCCGAGTCCAGACCTGTGACACAGACAGCC 177  
 Db |||||||  
 QY 194 TTCTGGAGTCAGGGGTGACAGACCTCCGAGTCCAGACCTGTGACACAGACAGCC 253  
 Db |||||||  
 QY 178 TCCTTAATGACACGCTCCATGTACGACCTGTAACTATACGCCATGCTTATACGT 237  
 Db |||||||  
 QY 254 TCCTTAATGACACGCTCCATGTACGACCTGTAACTATACGCCATGCTTATACGT 313  
 Db |||||||  
 QY 238 AACTTGTACTGACGTCAC---GATGGGTGTAACGCTGCTGCTTATGATTCATAGTG 294  
 Db |||||||  
 QY 314 AACTTATATGACGTCACGATGATGATGACGTCGCTGCTTATGATTCATAGTG 373  
 Db |||||||  
 QY 295 AGTTCTTAATAATACGACGTCGACCGCTTCTGAGGCTTGAACAGACTGCTGCTG 354  
 Db |||||||  
 QY 374 AGTTTCTTAATAAACGACGACGACGACTTCTTGGAGCTTGAACAGACTGCTGCTG 433  
 Db |||||||  
 QY 355 TGTTCCTTACGACGATTTCTGTTCTGATGCTGCTGCTGCTGCTGCTGCTGCTG 414  
 Db |||||||  
 QY 434 TGTTCCTTACGACGATTTCTGTTCTGATGCTGCTGCTGCTGCTGCTGCTGCTG 491  
 Db |||||||  
 QY 415 ACACGACGTCACGACG---AGACAGACGATTTTACTATATGAAAGTGAATTA 470  
 Db |||||||  
 QY 492 ACACGACGTCACGACG---AGACAGACGATTTTACTATATGAAAGTGAATTA 551  
 Db |||||||  
 QY 471 ATTACACTACATTTGACATGAGTATTTCACTGATTAAGTGTACAGGCTTAA 526  
 Db |||||||  
 QY 552 CTTTACACTACATTTGACATGAGTATTTCACTGATTAAGTGTACAGGCTTAA 607  
 Db |||||||  
 RESULT 8  
 AL583884 202342 bp DNA linear ROD 24-DEC-2002  
 LOCUS Mouse DNA sequence from clone RP23-321B16 on chromosome 15,  
 DEFINITION complete sequence.  
 ACCESSION AL583884  
 VERSION AL583884.20 GI:27368255  
 KEYWORDS HTG.  
 SOURCE Mus musculus (house mouse)  
 ORGANISM Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.  
 1 (bases 1 to 202342)  
 DIRECT SUBMISSION  
 Submitted (21-DEC-2002) Wellcome Trust Sanger Institute, Hinxton,  
 Cambridgeshire, CB10 1SA, UK. E-mail enquiries:  
 humquerry@sanger.ac.uk Clone requests: clonerequests@sanger.ac.uk  
 On Dec 23, 2002 this sequence version replaced g1:26788018.  
 Sequence from the Mouse Genome Sequencing Consortium whole genome  
 shotgun may have been used to confirm this sequence. Sequence data  
 from the whole genome shotgun alone has only been used where it has  
 a phred quality of at least 30.  
 ----- Genome Center  
 Center: Wellcome Trust Sanger Institute  
 Center code: SC  
 Web site: http://www.sanger.ac.uk  
 Contact: humquerry@sanger.ac.uk  
 -----  
 During sequence assembly data is compared from overlapping clones.  
 Where differences are found these are annotated as variations  
 together with a note of the overlapping clone name. Note that the

variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.

This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest, except on the rare occasion of the clone being a YAC.

The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em., EMBL; Sw., SWISSPROT; Tr., TREMBL; Wp., WORMPEP; Information on the WORMPEP database can be found at [http://www.sanger.ac.uk/Projects/C\\_elegans/wormpep](http://www.sanger.ac.uk/Projects/C_elegans/wormpep) RP23-324B16 is from the RPCI-23 Mouse PAC Library constructed by the group of Pieter de Jong. For further details see <http://www.chori.org/bacpac/home.htm>

## FEATURES

Source  
Location/Qualifiers  
1..202342  
/organism="Mus musculus"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:10090"  
/chromosome="15"  
/clone="RP23-324B16"  
/clone\_11b="RPCI-23"  
BASE COUNT 60264 a 39265 c 39460 g 63353 t  
ORIGIN

Query Match 39.6%; Score 403.2; DB 10; Length 202342;  
Best Local Similarity 89.7%; Pred. No. 1.9e-108;

Matches 481; Conservative 0; Mismatches 43; Indels 12; Gaps 4;

QY 1 CCCCTATTCCTGCTCAGATTAAGATTGCCAAATACCTTGTAAGTAACTGGCTGTG 60  
Db 184518 CCCCTATTCCTGCTCAGATTAAGATTGCCAAATACCTTGTAAGTAACTGGCTGTG 184517  
QY 61 CTGGAACAGCTAAGCACTAAGCTGTGAGAGA---CTTTGCTCTTAAGAAAGACTGCAGC 117  
Db 184578 CTGGAACAGCTAAGCACTAAGCTGTGAGAGA---CTTTGCTCTTAAGAAAGACTGCAGC 184637  
QY 118 TTTGAGGCTCAGGGGTGAGAGACCCCTCCGTAGCTCCGAGACCGTGTGACAGCAGCAGCC 177  
Db 184638 TTTGAGGCTCAGGGGTGAGAGACCCCTCCGTAGCTCCGAGAGAGTGTGACAGCAGCAGCC 184697  
QY 178 TCCCTTAATGACACGCTGCCATGTACGACACCTGTAACTTATCAGCCATGCTCATTTAGCT 237  
Db 184698 TCCCTTAATGACACGCTGCCATGTACGACACCTGTAACTTATCAGCCATGCTCATTTAGCT 184757  
QY 238 AACCTTTGACCTGTCAGTAC---GATGGGTGTAAACAGCTGCTCTTTGATTCTATAGTG 294  
Db 184758 AACCTTTGACCTGTCAGTAC---GATGGGTGTAAACAGCTGCTCTTTGATTCTATAGTG 184817  
QY 295 AGTTCTTAAATACAGCTGACCGGCTTTCGACGGCTTGAAGAAGACCTCGGCTCTG 354  
Db 184818 AGTTCTTAAATACAGCTGACCGGCTTTCGACAGACTTGAAGAAGACCTCGGCTCTG 184877  
QY 355 TGGTGCTCTTAACGAAGTATTCGTTCTAGTCTGCGGTGTGCTGGGTGAGAGTGTTGAA 414  
Db 184878 TGGTGCTCTCTATGAAGTATTCGTTCTAGTCTGCGGTGTGCTGGGTGAGAGTGTTGAA 184935  
QY 415 ACAGGAGCTCATCAAGG---AGACAGACAGTATTTTGGACTAATATGACATAGAGATTA 470  
Db 184936 ACAGGAGCTCATCAAGG---AGACAGACAGTATTTTGGACTAATATGACATAGAGATTA 184995  
QY 471 ATTACACTACATTTGATGAGATTAATTCAGTGAATAAAGTGTGCGGGTAA 526  
Db 184996 CTTTACACTACATTTGATGAGATTAATTCAGTGAATAAAGTGTGCGGGTAA 185051

RESULT 9

AC126556  
LOCUS AC126556 179685 bp DNA linear ROD 11-JUN-2003  
DEFINITION Mus musculus chromosome 15 clone RP23-64L3, complete sequence.  
AC126556  
VERSION AC126556.4 GI:11581824  
KEYWORDS hmc.  
SOURCE Mus musculus (house mouse)  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 179685)

Wilson,R.K.

TITLE The sequence of Mus musculus clone

Unpublished

JOURNAL 2 (bases 1 to 179685)

REFERENCE McPherson,J.D. and Waterston,R.H.

AUTHORS McPherson,J.D. and Waterston,R.H.

TITLE Direct Submission

JOURNAL Submitted (06-JUL-2002) Genome Sequencing Center, 4444 Forest Park Parkway, St. Louis, MO 63108, USA

REFERENCE 3 (bases 1 to 179685)

McPherson,J.D. and Waterston,R.H.

TITLE Direct Submission

JOURNAL Submitted (07-JAN-2003) Genome Sequencing Center, 4444 Forest Park Parkway, St. Louis, MO 63108, USA

REFERENCE 4 (bases 1 to 179685)

Wilson,R.K.

TITLE Direct Submission

JOURNAL Submitted (11-JUN-2003) Genome Sequencing Center, 4444 Forest Park Parkway, St. Louis, MO 63108, USA

COMMENT On Jun 11, 2003 this sequence version replaced gi:27502524.

----- Genome Center -----  
Center: Washington University Genome Sequencing Center  
Center code: WUGSC  
Web site: <http://genome.wustl.edu>  
Contact: [submissions@watson.wustl.edu](mailto:submissions@watson.wustl.edu)  
----- Project Information -----  
Center project name: M\_BA0064L03  
-----

## FEATURES

Source  
Location/Qualifiers  
1..179685  
/organism="Mus musculus"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:10090"  
/chromosome="15"  
/clone="RP23-64L3"  
BASE COUNT 47438 a 39048 c 39639 g 53560 t  
ORIGIN

Query Match 33.2%; Score 337.2; DB 10; Length 179685;  
Best Local Similarity 84.5%; Pred. No. 8.9e-89;

Matches 443; Conservative 0; Mismatches 63; Indels 18; Gaps 5;

QY 16 AGATTAAAGATTGGCAAAATA-CCCTGGAAGTAAAGTGGCTGGAGTGAACAGCTAA 74  
Db 56268 AAAAAAGAAATTTGCCAAAATATACCTTCGAACTAAGTTGCTGTGCTGAGAAACCTTAA 56327  
QY 75 GCCTTAAGCTGT---TGAGAGACTTGTCTTAAAGAGACTGCAAGCTTGTG-----GC 125  
Db 56328 GCCTTAAGCTGTCTTGAAGAGACTTGTGCTGTGAAGAGACTGCAAGCTTGTGAGCCAGG 56387  
QY 126 TCAGGGGTGACAGACCCCTCCCTAGCTCCGAGACCGTGTGACACAGCAGCAGCTCTTAAT 185  
Db 56388 TCAGGGGTGACAGACCCCTCCCTAGCTCCGAGACCGTGTGACACAGCAGCAGCTCTTAAT 56447  
QY 186 GACAGCGCTCATGTAAGCAGCTGTAACTATACAGCCATGCTCATTAAGTAACTTTGT 245  
Db 56448 GACAGCGCTCATGTAAGCAGCTGTAACTATACAGCCATGCTCATTAAGTAACTTTGT 56507  
QY 246 ACTGTAGCTCAC---GATGGGTGTAAACAGCTGCTCTTTGATTCTCATAGTGAATCTCT 302  
Db 56508 ACTGTAGCTCACAGTGAATGAGTGTGACAGCTGCAAGCTTGTGATTCAATAGTGAATTTCT 56567

QY 303 AAAATACACGCTACCGGCTTTCGACGGCTTTGAACAGAACTCGCTCTGTGTGCT 362  
DB 56568 AAAAGACACGACCCAGCTTTCGACAGCTTTCGAACATCTGTGTGCTGTGCT 56627  
QY 363 CTAAAGATATCTGTCTCTACTGCTGGGTGTGCTGGGTGTGCTGTGCTGTGCT 422  
DB 56628 CTATGAGATATCTGTCTCTACTGCTGGGTGTGCTGGGTGTGCTGTGCTGTGCT 56687  
QY 423 TCATCAAGAGACAGACAGTATTTTGTACATATGATGATGATGATGATGATGAT 482  
DB 56688 ATCAAGAGACAGACAGATATTTTGTACATATGATGATGATGATGATGATGAT 56742  
QY 483 TTGTACATGAGATATTTCACTGATGATGATGATGATGATGATGATGATGAT 526  
DB 56743 TTGTACATGAGATATTTCACTGATGATGATGATGATGATGATGATGATGAT 56786

RESULT 10  
AX163752 1022 bp DNA linear PAT 22-JUN-2001  
LOCUS AX163752 Sequence 16 from Patent WO0138579.  
DEFINITION AX163752  
ACCESSION AX163752  
VERSION AX163752.1 GI:14544858  
KEYWORDS  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1  
AUTHORS Gould-Rothberg, B.E., Dipippo, V.A., Ramseh, T.M. and Gervelin, R.W.  
TITLE Method of identifying toxic agents using nsaid-induced differential  
JOURNAL gene expression in liver  
Curagen Corporation (US)  
16 31-MAY-2001;  
location/Qualifiers  
FEATURES  
source 1..1022  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
BASE COUNT 273 a 208 c 205 g 336 t  
ORIGIN

Query Match 28.6%; Score 290.6; DB 6; Length 1022;  
Best Local Similarity 64.5%; Pred. No. 4.5e-75;  
Matches 664; Conservative 0; Mismatches 304; Indels 61; Gaps 13;

QY 1 CCCCTATCTGCTGATTAAGATTCGCAAAATACCTGTGACATGTC---GTT 57  
DB 1 CCCCTATCTGCTGATTAAGATTCGCAAAATACCTGTGACATGTC---GTT 60  
QY 58 GTGCTGAGAACAGCTAAGCTAAGCTGTGAGAGACTTGTCTTAAAGACTGAC 117  
DB 61 GTGCTGAGAACAGCTAAGCTAAGCTGTGAGAGACTTGTCTTAAAGACTGAC 120  
QY 118 TTCTGGGCTCAGG-GGTGCAGACCTCCCGCTAGC-----TCCAGACCGTGTGACACA 169  
DB 121 TTCTGGGCTCAGG-GGTGCAGACCTCCCGCTAGC-----TCCAGACCGTGTGACACA 180  
QY 170 GCACAGCTCCTTAAGACAGCGCTGCATGTAAGCAACCTGTAACCTTATGACCCATGCT 229  
DB 181 GAACAGCTCCTCGAATGAAGCGTGTGCAATGAACAGCTGATGATGACAGCGTGT 240  
QY 230 CATTAAGTAACTTGTACTGTACGTACATGATGATGATGATGATGATGATGATGAT 289  
DB 241 CCGGACATTAACATGTACTGTATGATGATGATGATGATGATGATGATGATGATGAT 300  
QY 290 TACTGAGTCTCTAAATATACACCTGACCGCTTTCGACGGCTTGAACAGAACTGTGCG 349  
DB 301 TTGTGATTTTCAATAGGAGATTTTCTGACACGCTTTTGGGAGATTTGAACAGAACTG 358  
QY 350 TCTGTGTGCTGCTTAAAGATATCTGTCTCTAGTGTGCTGGGTGTGCTGGGTGTGCT 409  
DB 359 ---CTATTTCTCTAATGAAGAAATCTGT---TAGCTGTGCTGTGCTGGGTGTGCTGGGTGT 412

QY 410 GTGAAACACGACGCTACCAAGAGACAGACAGTATTTGAC-TAAATGATGATGATGAT 468  
DB 413 GT-----GTGATCAAGAGACAAAGACATATTTTTCACAAATACGAACTGAG-- 460  
QY 469 TAAATTAACATTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 525  
DB 461 -ATTTACATTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 518  
QY 526 AGCTTTTAACGATTAATTTCTGTCAAAAGCTGATGATGATGATGATGATGATGATGAT 585  
DB 519 ATTTCTGTCAAAATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 578  
QY 586 TCTCTGTGAGACCCCTCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 645  
DB 579 TAGCTTTTCTGCTTCTGCTTACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 638  
QY 646 TTAACAGCAATATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 705  
DB 639 TTAACAGCAATATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 698  
QY 706 TTAACAGCAATATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 765  
DB 699 TGACTGAGCAATATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 757  
QY 766 TTTTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 825  
DB 758 TTTTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 812  
QY 826 GCTTCCCTGCTGCTTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 885  
DB 813 GCTTCCCTGCTGCTTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 872  
QY 886 TAAACATTTTAAATA-----TTGAAACATGATGATGATGATGATGATGATGATGAT 928  
DB 873 TAAACATTTTAAATA-----TTGAAACATGATGATGATGATGATGATGATGATGAT 932  
QY 929 TAAACATTTTAAATA-----TTGAAACATGATGATGATGATGATGATGATGATGAT 988  
DB 933 TAAACATTTTAAATA-----TTGAAACATGATGATGATGATGATGATGATGATGAT 991  
QY 989 CTTCGCTGA 997  
DB 992 CTTCGCTGA 1000

RESULT 11  
BC050687 2302 bp mRNA linear PRI 11-APR-2003  
LOCUS BC050687  
DEFINITION Homo sapiens, ras-related G3 botulinum toxin substrate 1 (rho  
family, small GTP binding protein Rac1), clone MGC:60264  
IMAGE:6149377, mRNA, complete cds.  
ACCESSION BC050687  
VERSION BC050687.1 GI:29792301  
KEYWORDS Homo sapiens (human)  
SOURCE Homo sapiens  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1  
AUTHORS Strausberg, R.  
TITLE Direct Submission  
JOURNAL Submitted (08-APR-2003) National Institutes of Health, Mammalian  
Gene Collection (MGC), Cancer Genomics Office, National Cancer  
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,  
USA  
REMARK NIH-MGC Project URL: <http://mgc.ncl.nih.gov>  
COMMENT Contact: MGC help desk  
Email: [cgabs-remail.nih.gov](mailto:cgabs-remail.nih.gov)  
Tissue Procurement: ATCC/DCID/DTP  
CDNA Library Preparation: Life Technologies, Inc.  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LIML)  
DNA sequencing by: Sequencing Group at the Stanford Human Genome

Center, Stanford University School of Medicine, Stanford, CA 94305  
 Web site: <http://www-sngc.stanford.edu>  
 Contact: (Dickson, Mark) mcd@paxil.stanford.edu  
 Dickson, M., Schmutz, J., Grimwood, J., Rodriguez, A., and Myers, R. M.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/BLN at: <http://image.llnl.gov>  
 Series: IRAC Plate: 110 Row: K Column: 14  
 This clone was selected for full length sequencing because it passed the following selection criteria: matched mRNA g1: 9845510.

## FEATURES

source

location/Qualifiers

## CDS

1. 2302  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="MGC:60264 IMAGE:6149377"  
 /tissue\_type="Skin, melanotic melanoma."  
 /lab\_host="DH10B"  
 /note="Vector: pCMV-SPORT6"  
 189..767  
 /codon\_start=1  
 /product="ras-related C3 botulinum toxin substrate 1 (rho family, small GTP binding protein Rac1)"  
 /protein\_id="AAH50687.1"  
 /db\_xref="GI:29792302"  
 /db\_xref="LocusID:5879"  
 /translation="MOAICVAVGVGAVGKCTLLISYTNAPGEYIPVVDNSANYMDGPVNLGMDTAGEYDRLRPLSPOTDFVLCISLSPSEFVRAKWTPEVHHCNPPIIVTGTLDRDDKDTLEKKEKLTPTITPDGLAMKEIGAVVLECSALTQRLKTYFDIAIRAVLCPPVKKRKRCLL"

BASE COUNT 617 a 524 c 498 g 663 t  
 ORIGIN

Query Match 28.6%; Score 290.6; DB 9; Length 2302;  
 Best Local Similarity 64.5%; Pred. No. 4.8e-75;

Matches 664; Conservative 0; Mismatches 304; Indels 61; Gaps 13;

1 CCCCTATTCTTCGTCAGATTAAAGAAATGCCAAATACCTTGTAAGTTGC--GTT 57  
 1050 CCCCATTCTTGTCAGATTAAAGAGTCCAAATACCTTGTAAGTTGCATGTT 1109  
 58 GTGCGAAGACAGTAAAGCAGTAAAGCTTTGAGACATTTGGCTTAAGACGACG 117  
 1110 GTGCGAAGACAGCAGGACGATTTGCAAGACCTTCGCTTGAAGAGCGGTAGC 1169  
 118 TTCTGGGCTCAGG-GGTGACAGCCCTCCCGTAGC-----TCCGACACGCTGTGACACA 169  
 1170 TTCTGCACTTGAAGAGTGCAGACACTTGTCTCCATGATGATTCGACATGGTAAACGA 1229  
 170 GCACAGCCTCCTTAATGACAGCGTGCATGTACGACCTGTAATTCAGCCCATGCT 229  
 1230 GAACAGCCTCCGATGAGAGGTGTCGATTAACACACAGAGTGAAGACAGTGT 1289  
 230 CATTCAGTACTTGTACTGTACGTCAGATGGGTGAACAGCTGCTGCTTTGATTTCA 289  
 1290 CCCCACATTAACATGTACTGTGAAGAGTGAAGCGTGAAGCTCTTTGGATCACTC 1349  
 290 TAGTAGTCTCTTAATAATCAGAGTGAAGCGCTTCGAGAGCTTTGAACAGAACTGTGC 349  
 1350 TTTTGATTTGATACGAGTTTTCGACAGCTTTTGGGAGATTTTGAACAGAACTG-- 1407  
 350 TCCGTGTTCCTTAACGAAGTATTTGTTCTAGTGGTGGTGTGCTGGGTGAGTGT 409  
 1408 ---CTATTCTCTTAATGAAGAAATTCGTT--TAGCTGTGGGTGTGCGGGGTGGGGT 1461  
 410 GTGAAGACGAGCTATCAAGAGACAGAGTATTTTGC--TAATATGAAGTGAAGAT 468  
 1462 GT-----GTATCAAGAGACAAAGACGATTTTGTGAAATAATACGAAGTGAAG-- 1509  
 469 TAATTACATACATTTGATGAGTAA--TTCACTGAATTAAGTGTACGGGTAA 525

1510 --ATTTCATACATTTGATACAGAAATGAAGTCTACGGGTAAAACTTAAGTTA 1567  
 526 AGCTTTTAAACGTTAAATTTCTGTCAAAACAGATAGATGACAATGGCGATTTATCAGTG 585  
 1568 ATTCTGTCAAAATGACAGATAGATGATGAAGAAAGATGGATATATTCAGGAATGTCTT 1627  
 586 TCTCTTTGAGCCCCCTTCCCTGCTGTCCTCCCTCCAGATGGGGCTTGATCCATAT 645  
 1628 TAACTTTCTCTCTCTTACACCTGCGCATGCCCAAAATGGCATTTAAATTCATCT 1687  
 646 TTAACAGGCGCATCTGACATTTCTAACCTTACAGAGGCTTTCTTAGAGACCCCTTC 705  
 1688 TTAACAGTGTGTTCTGTGTAGCTTACCTTACAGAGGCTTTCTTAGAGACCCCTTC 1747  
 706 TTAACAGCAATATGTCTGACCTGTACATATGAATCTTCTGATATGATTCGAGAT 765  
 1748 TGACGAGCAATATGCGCT--CTGTGATATATTAATCTTTGATATGATTCAGAT 1806  
 766 TTTTGTGATATGATGAAGGCGCTTCTGTTTACACCTTCTTACTAGCTGACATAGT 825  
 1807 TTTTGTGATATGATGAAGGCGCTTCTGTTTACACCTTCTTACTAGCTGACATAGT 1861  
 826 GCTTCCCTCTGTTTCTGATTAACCTGGGTGAGAAATCAGCTGCGGCTTTACAGTTT 885  
 1862 GCTTCCCTTACTTTTCTGATTAACCTGGGTGAGAAATCAGCTGCGGCTTTACAGTTT 1921  
 886 TAACTATTTTGAATA-----TTCTGAACATCACTGTCTTCCAGAG 928  
 1922 TAAATATATTTAGATATTTCTTAAACATATGACCTTCTTACATCACTGTCTTCCAGAT 1981  
 929 TACCAACCTGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 988  
 1982 TACGACACCTGTCACTGTGACCAATATG-ACCCTCTTAACTCCGCCACGCGGACACAG 2040  
 989 CTTCGGGTA 997  
 2041 CCTCTGTGA 2049

DB 2041 CCTCTGTGA 2049

RESULT 12  
 AK054993  
 LOCUS  
 DEFINITION  
 Homo sapiens cDNA FLJ30431 fls, clone BRAC2008966, highly similar to RAS-RELATED C3 BOTULINUM TOXIN SUBSTRATE 1.  
 ACCESSION  
 AK054993.1 GI:16549633  
 VERSION  
 AK054993.1  
 KEYWORDS  
 oligo cloning; fls (full insert sequence).  
 SOURCE  
 Homo sapiens  
 ORGANISM  
 Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE  
 1  
 Tashiro, H., Yamazaki, M., Watanabe, K., Kumagai, A., Itakura, S., Fukuzumi, Y., Fujimori, Y., Komiyama, M., Sugiyama, T., Irie, R., Otsuki, T., Sato, H., Wakematsu, A., Ishii, S., Yamamoto, J., Igono, Y., Kawai-Hio, Y., Saito, K., Nishikawa, T., Kimura, K., Yamashita, H., Matsuo, K., Nakamura, Y., Sekine, M., Kikuchi, H., Kanda, K., Wagahtsuma, M., Murakawa, K., Kanehori, K., Takahashi, Fujii, A., Oshima, A., Sugiyama, A., Kawakami, B., Suzuki, Y., Sugano, S., Negahara, K., Masuno, Y., Nagai, K. and Isozaki, T.  
 NEDO human cDNA sequencing project  
 TITLE  
 NEDO human cDNA sequencing project  
 JOURNAL  
 Unpublished  
 2 (bases 1 to 2315)  
 Isozaki, T., Otsuki, T. and Sugiyama, T.  
 Direct Submission  
 Submitted (24-OCT-2001) Takara Isogen, Helix Research Institute, Genomics Laboratory, 1532-3 Yano, Kisarazu, Chiba 297-0812, Japan (E-mail: genomics@helix.co.jp, tel: 01-438-52-3975, fax: 01-438-52-3986)  
 NEDO human cDNA sequencing project supported by Ministry of Economy, Trade and Industry of Japan, cDNA full insert sequencing: Research Association for Biotechnology (RAB); cDNA library construction: Helix Research Institute (HRI) (supported by Japan Key Technology Center etc.); 5'- 6 3'-end one pass sequencing: RAB.

HRI, and Biotechnology Center, National Institute of Technology and Evaluation; clone selection for full insert sequencing; RAB and HRI.

FEATURES  
source  
1..2315  
Location/Qualifiers  
/organism="Homo sapiens"  
/mol\_type="mrna"  
/db\_xref="taxon:9606"  
/clone="BRACE2008968"  
/issue\_type="cerebellum"  
/clone\_1ib="BRACE2"  
/note="cloning vector: pHE18SFL3"

BASE COUNT 603 a 531 c 501 g 680 t  
ORIGIN

Query Match 28.6%; Score 290.6; DB 9; Length 2315;  
Best Local Similarity 64.5%; Pred. No. 4.8e-75;  
Matches 664; Conservative 0; Mismatches 304; Indels 61; Gaps 13;

1 CCCCTATTCCTGCTCAGATTAGAAATGCCAAATACCTTGTGAAGTTCGCTGTT 57  
1078 CCCCATTCCTGCTCAGATTAGAAATGCCAAATACCTTGTGAAGTTCGCTGTT 1137  
58 GTCTGAGAACACGTAGACCTAAGCTGTGAGAGACTTGTCTTAAAGAACTGCAC 117  
1138 GTCCGAGAACACCGGACCTGACCTTGCAGAAAGACCTTGTCTTGAAGAGCGGTAC 1197  
118 TTTCTGGGCTCAGG-GGTGAGACCTCCCGTAGC-----TCCGAGACCGTGTACACA 169  
1138 TTTCTCAGTAGTAGAGGTGCAGACCTGTCTCTCATATGATTCAGATGCGTAGACA 1257  
170 GCACAGCCTCTTAAATGACACGCTGCATGACGACCTGTAACCTATACGCCATGCT 229  
1258 GAAACAGCCTCCGAGATGAAAGCCTTGCCATTAACCTACACAGTAGTAGACAGCTGT 1317  
230 CATTCAGTACTTGTACTGTACGTACGATGAGGTGTAAACAGCTGTCTCTTGAATTTCA 289  
1318 CCCGACATTAACATTTACTGTATGAGAGTACGACGTCAGCTCTTGTGATCAGTTC 1377  
290 TAGTAGTCTCTTAAATACACAGCTGACCGGCTTCGAGGCTTGAACAGAACTGTGC 349  
1378 TTTTGATTTTCATACGAGATTTTCTGACCAAGCTTTTGCAGATTTTGAACAGAACTG-- 1435  
350 TCCCTGTGCTCTTAACGAACTATTCCTGCTGAGTGTGCTGTGCTGTGAGTGT 409  
1436 ----CTATTCCTCTAATGAAATCTGT--TAGCTGTGGGTGTGCCGGTGGGTGT 1489  
410 GTGAACACGAGCTCATCAAGAGACAGACAGTATTTTGTAC-TAATATGAAGTAGACAT 468  
1490 GT-----GTATCAAAAGACAAAGACAGTATTTTGAACAAATACGAAGTGTGAG-- 1537  
469 TAATTTACACTATTTGATGATGAGTAA--TTCACTGATTAAGTGTACAGGTTAA 525  
1538 --ATTACACTATCTATGTAAGAGATGAAGATGTACACGGTAAAGAACTCTAAAGGTAA 1595  
526 AGCTTTTAAACGTTAATTTCTGTCAACAGTAGATGACAAATGCGCCGATTTATCAGTG 585  
1596 ATTCTGTCAAAATGAGTAGATGAGTAAAGAAAGGTGTGTATTAACAGAAATGTTTCT 1655  
586 TCTCTCTGAGCCCTTCCCTTCCCTGTCTCCCTCCCAAGATGGGCGGTGATGCCATAT 645  
1656 TAAGCTTTCCTTCTCTTACACCTGCTCCATGCTCCCAAAATGGGCATTTAATCATCT 1715  
646 TTAACCTGGCCATCTCAAGTTGCTAAGTGAAGAAAGTCTTTCTTTGAGACCCCTTTC 705  
1716 TTAACCTGGTCTTCTGTTAGTCCCTAAGTGAAGTGTCTTTCTTATAGAACCCCTTTC 1775  
706 TTAACGAGCAATATCTGACCTGTACTAATAGATCTTCTGATATGATTCGAGATTT 765  
1776 TGACGTAGCAATATGAGCT-CTTGTATATATAAATCTTCTGATATGATTAAGAGTT 1834  
766 TTTTGTGATAGTAGTAGAAGTGTCTCTGTTTTCACCTCTTACTACAGCTGACGTACT 825

DB 1835 TTTTGTGATAGTAGTAGAAGTGTCTTCCATGTAC-----TTATATCAGAGCTAATAGT 1889  
QY 826 GCTTCCCTGCTTCTTCTAGTACTGGCTGTAGAAATCAGTGTCTGCGCTTACGTTT 885  
DB 1890 GCTTCTCTAGTGTCTTCTAGTACTGAGTAAATCAATGATGTGACGCTTATAGTTT 1949  
QY 886 TAACATATTTTATGATA-----TTCTGAAACATCAGCTGTTCGACAG 928  
DB 1950 TAAATATTTTATGATATATCTTAAACTATGAACTCTTACATCATCTCTTCCAGAT 2009  
QY 929 TACCAACTGTCTGATGATGATGCGCCCTCTAGACCTACCCAGCGGACACATG 988  
DB 2010 TACCGACACTGTCTGATGATGATGCGCCCTCTAGACCTACCCAGCGGACACAG 2068  
QY 989 CTTCGGCTA 997  
DB 2069 CCTCTGTGA 2077

RESULT 13  
HSA132695  
LOCUS HSA132695 28567 bp DNA linear PRI 09-NOV-2000  
DEFINITION Homo sapiens rac1 gene.  
ACCESSION AJ132695  
VERSION AJ132695.5 GI:8574037  
KEYWORDS alternative splicing; Alu; AluJo; AluSg; AluSp; AluX; AT-rich; CT-rich; MIR; rac1 gene; rac1 protein; Rac1 protein; repetitive sequence.

SOURCE  
ORGANISM Homo sapiens (human)

REFERENCE  
AUTHORS Matos, P., Skaug, J., Marques, B., Beck, S., Verislimo, F., Geppach, C., Boavida, M.G., Scherer, S.W. and Jordan, P.  
TITLE Small GTPase Rac1: structure, localization, and expression of the human gene

JOURNAL BIOchem. Biophys. Res. Commun. 277 (3), 741-751 (2000)  
MEDLINE 20517245  
PUBMED 11062023  
REFERENCE 2 (bases 1 to 28567)  
AUTHORS Jordan, P.  
TITLE Direct Submission

JOURNAL Submitted (03-FEB-1999) Jordan P., Centro de Genetica Humana, Laboratorio de oncobiologia, Instituto Nacional de Saude, Dr. Ricardo Jorge, Avenida Padre Cruz, 1649-016 Lisboa, PORTUGAL  
COMMENT On Jun 20, 2000 this sequence version replaced gi:7248282.  
Related sequence AI638561.

FEATURES  
source  
1..28567  
Location/Qualifiers  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
/chromosome="7"  
/map="7p22"  
/clone="H. NH0425A05, RFI 11"

repeat\_region 1..109  
/rpt\_family="AluSg"  
repeat\_region 123..422  
/rpt\_family="AluSc"  
repeat\_region 441..742  
/rpt\_family="AluSg"  
gene 743..28117  
/gene="rac1"  
promoter 743..1226  
/gene="rac1"  
repeat\_region 911..983  
/rpt\_family="GC-rich"  
misc\_feature 936  
/gene="rac1"  
misc\_feature 1026  
/gene="rac1"

misc\_feature  
1026  
/gene="rac1"



```
/note="alternative transcription start site"
/misc_feature 1037
/gene="rac1"
/note="alternative transcription start site"
/misc_feature 1053
/gene="rac1"
/note="alternative transcription start site"
repeat_region 1106..1175
/rpt_family="GC-rich"
join(1227..1261,13701..13772,18413..18530,22804..22860,
24268..24330,26010..26169,26458..26588)
/gene="rac1"
/note="alternative"
/codon_start=1
/product="Rac1 protein"
/protein_id="CA10733.6"
/db_xref="GI:8574039"
/translation="MOAICVVGDAVGKTCLLISYTTNAPGEYIPTYVDNYSANV
MWDKPYNLGMDYAGODYDRKPLSTPQVGHYGDITSRGDKPLADVFLIFS
LVSPASENVAKWPPEVRHHCPTPIIVGTGLDLDDKDTIEKLEKLTPTTYPQ
GLAMAKEIGAVKYLECSALTQRIKTYDEAIRAVLPPVKKRRKRCCLL"
join(1227..1261,13701..13772,18413..18530,24268..24330,
26010..26169,26458..26588)
/gene="rac1"
/codon_start=1
/product="Rac1 protein"
/protein_id="CAB53579.5"
/db_xref="GI:8574038"
/translation="MOAICVVGDAVGKTCLLISYTTNAPGEYIPTYVDNYSANV
MWDKPYNLGMDYAGODYDRKPLSTPQVGHYGDITSRGDKPLADVFLIFS
LVSPASENVAKWPPEVRHHCPTPIIVGTGLDLDDKDTIEKLEKLTPTTYPQ
GLAMAKEIGAVKYLECSALTQRIKTYDEAIRAVLPPVKKRRKRCCLL"
HHCNPPIIVGTGLDLRDDKDTIEKLEKLTPTTYPQGLAMAKEIGAVKYLECSAL
TORGIKTVEDEAIRAVLCPPVKKRRKRCCLL"
<1227..1261
/gene="rac1"
/number=1
1262..13700
/gene="rac1"
/number=1
1390..1419
/rpt_family="GC-rich"
repeat_region 1752..1793
/rpt_family="FLAM"
repeat_region 1794..2005
/rpt_family="Alu"
repeat_region 2140..2171
/rpt_family="Simple_repeat"
repeat_region 2666..2972
/rpt_family="Alu"
repeat_region 2980..3267
/rpt_family="Alu"
repeat_region 3268..3289
/rpt_family="Alu"
repeat_region 3802..4102
/rpt_family="Alu"
repeat_region 4123..4386
/rpt_family="Alu"
repeat_region 4538..4719
/rpt_family="L1"
repeat_region 4754..5053
/rpt_family="L1"
repeat_region 5297..5594
/rpt_family="Alu"
repeat_region 5635..5757
/rpt_family="Alu"
repeat_region 5758..6062
/rpt_family="Alu"
repeat_region 6063..6198
/rpt_family="Alu"
repeat_region 6222..6401
/rpt_family="Alu"
repeat_region 7162..7238
/rpt_family="L1"
repeat_region 7348..7367
/rpt_family="Simple_repeat"
repeat_region 7368..7656
/rpt_family="Alu"
repeat_region 7880..8183
/rpt_family="Alu"
repeat_region 8484..8778
/rpt_family="Alu"
repeat_region 9071..9365
/rpt_family="Alu"
repeat_region 10279..10583
/rpt_family="Alu"
repeat_region 10609..10864
/rpt_family="Alu"
repeat_region 10970..11276
/rpt_family="L1"
repeat_region 11276..11594
/rpt_family="Alu"
repeat_region 11744..11825
/rpt_family="L1"
repeat_region 11764..11987
/rpt_family="MIR"
repeat_region 12989..13307
/rpt_family="Alu"
repeat_region 13701..13772
/gene="rac1"
/number=2
13773..18412
/gene="rac1"
/number=2
13954..14235
/rpt_family="Alu"
repeat_region 14807..14831
/rpt_family="Simple_repeat"
repeat_region 14836..15128
/rpt_family="Alu"
repeat_region 15408..15715
/rpt_family="Alu"
repeat_region 15872..15896
/rpt_family="AT-rich"
repeat_region 15902..16186
/rpt_family="Alu"
repeat_region 16187..16314
/rpt_family="Alu"
repeat_region 16413..16585
/rpt_family="Simple_repeat"
repeat_region 16586..16679
/rpt_family="FLAM"
repeat_region 16707..16810
/rpt_family="L1"
repeat_region 16811..17108
/rpt_family="Alu"
repeat_region 17109..17345
/rpt_family="L1"
repeat_region 17377..17686
/rpt_family="Alu"
repeat_region 17703..17729
/rpt_family="Simple_repeat"
repeat_region 18413..18530
/gene="rac1"
/number=3
18531..22803
/gene="rac1"
/number=3
19287..19436
/rpt_family="MIR"
repeat_region 21161..21184
/rpt_family="Simple_repeat"
repeat_region 21185..21351
/rpt_family="Alu"
repeat_region 21365..21403
/rpt_family="Alu"
misc_difference 21364^21365
```

```
repeat_region 7348..7367
/rpt_family="Simple_repeat"
repeat_region 7368..7656
/rpt_family="Alu"
repeat_region 7880..8183
/rpt_family="Alu"
repeat_region 8484..8778
/rpt_family="Alu"
repeat_region 9071..9365
/rpt_family="Alu"
repeat_region 10279..10583
/rpt_family="Alu"
repeat_region 10609..10864
/rpt_family="Alu"
repeat_region 10970..11276
/rpt_family="L1"
repeat_region 11276..11594
/rpt_family="Alu"
repeat_region 11744..11825
/rpt_family="L1"
repeat_region 11764..11987
/rpt_family="MIR"
repeat_region 12989..13307
/rpt_family="Alu"
repeat_region 13701..13772
/gene="rac1"
/number=2
13773..18412
/gene="rac1"
/number=2
13954..14235
/rpt_family="Alu"
repeat_region 14807..14831
/rpt_family="Simple_repeat"
repeat_region 14836..15128
/rpt_family="Alu"
repeat_region 15408..15715
/rpt_family="Alu"
repeat_region 15872..15896
/rpt_family="AT-rich"
repeat_region 15902..16186
/rpt_family="Alu"
repeat_region 16187..16314
/rpt_family="Alu"
repeat_region 16413..16585
/rpt_family="Simple_repeat"
repeat_region 16586..16679
/rpt_family="FLAM"
repeat_region 16707..16810
/rpt_family="L1"
repeat_region 16811..17108
/rpt_family="Alu"
repeat_region 17109..17345
/rpt_family="L1"
repeat_region 17377..17686
/rpt_family="Alu"
repeat_region 17703..17729
/rpt_family="Simple_repeat"
repeat_region 18413..18530
/gene="rac1"
/number=3
18531..22803
/gene="rac1"
/number=3
19287..19436
/rpt_family="MIR"
repeat_region 21161..21184
/rpt_family="Simple_repeat"
repeat_region 21185..21351
/rpt_family="Alu"
repeat_region 21365..21403
/rpt_family="Alu"
misc_difference 21364^21365
```



University Genome Sequencing Center. For additional information about the map position of this sequence, see <http://www.hbgrl.nih.gov/DIR/GRB/CHR7>, send mailto:egreen@hbgrl.nih.gov, or see <http://genome.wustl.edu/gsc>

## SOURCE INFORMATION:

The RPCT-11 human BAC library was made from the blood of one male donor, as described by Osogawa, K., Woon, P.Y., Zhao, B., Frengen, E., Tateno, M., Cataneese, J.J. and de Jong, P.J. (1998) An improved approach for construction of bacterial artificial chromosome libraries. Genomics 51:1-8. The clone may be obtained either from Research Genetics, Inc. (<http://www.resgen.com>) or Pieter de Jong and coworkers at the Roswell Park Cancer Institute (<http://pacpac.med.buffalo.edu>)

VECTOR: pBAC3.6

## NEIGHBORING SEQUENCE INFORMATION:

The clone sequenced to the right is CTD-2195F2, 200 bp overlap. Actual start of this clone is at base position 1 of RP11-425P5; actual end is at base position 212633 of RP11-425P5.

## Location/Qualifiers

```
FEATURES
Source
1. .212827
   /organism="Homo sapiens"
   /mol_type="genomic DNA"
   /db_xref="taxon:9606"
   /chromosome="7"
   /map="7"
   /clone="RP11-425P5"
   /clone_id="RPCT-11"
   repeat_region
   1623. .1696
   /rc_family="MIR"
   1698. .2006
   /rc_family="Alu"
   2024. .2337
   /rc_family="Alu"
   repeat_region
   3006. .3093
   /rc_family="L1"
   3453. .3552
   /rc_family="L2"
   4924. .5046
   /rc_family="ERV1"
   5318. .5693
   /note="similar to EST BF996284 (NID:g12402607)"
   repeat_region
   5580. .5845
   /rc_family="Alu"
   5925. .6277
   /rc_family="MALR"
   6842. .7041
   /note="similar to EST BF369507 (NID:g11331532)"
   repeat_region
   7364. .7661
   /rc_family="Alu"
   8029. .8048
   /rc_family="(CAAA)n"
   8379. .8678
   /rc_family="Alu"
   8560. .8762
   /note="similar to EST AA699398 (NID:g2702592) z140a03.s1"
   repeat_region
   8720. .9031
   /rc_family="Alu"
   9083. .9560
   /note="similar to EST AA633603 (NID:g2556817) ae66b02.s1"
   repeat_region
   9598. .9899
   /rc_family="Alu"
   9871. .10098
   /note="similar to EST AA699398 (NID:g2702592) z140a03.s1"
   repeat_region
   10746. .11044
   /rc_family="L1"
   11075. .11101
   /rc_family="AT-rich"
   repeat_region
   11102. .11372
   /rc_family="Alu"
   11411. .11465
   /rc_family="CA-rich"
   repeat_region
   11552. .11644
```

```
repeat_region
11685. .12193
/rc_family="L2"
repeat_region
12194. .12486
/rc_family="L1"
repeat_region
12487. .13003
/rc_family="Alu"
repeat_region
13004. .13305
/rc_family="L1"
repeat_region
13306. .13501
/rc_family="Alu"
repeat_region
13503. .13892
/rc_family="L1"
misc_feature
13893. .14224
/note="similar to EST BE061102 (NID:g8405752)"
repeat_region
14244. .14544
/rc_family="Alu"
repeat_region
14546. .14828
/rc_family="Alu"
repeat_region
15091. .15189
/rc_family="MIR"
repeat_region
15672. .16449
/rc_family="MIR"
repeat_region
16640. .16844
/rc_family="Acho"
repeat_region
16917. .16959
/rc_family="Acho"
repeat_region
16960. .17287
/rc_family="Alu"
repeat_region
17288. .17572
/rc_family="Acho"
misc_feature
17610. .17938
/note="similar to EST BF996427 (NID:g12402750)"
repeat_region
17793. .18068
/rc_family="Alu"
repeat_region
18251. .18537
/rc_family="Alu"
misc_feature
18485. .18884
/note="similar to EST BE062347 (NID:g8406997)"
repeat_region
18661. .18809
/note="similar to EST BG186544 (NID:g13708231)"
repeat_region
18957. .19160
/rc_family="L1"
repeat_region
19314. .19528
/rc_family="L1"
repeat_region
19819. .20135
/rc_family="Alu"
repeat_region
21101. .21269
/rc_family="L1"
repeat_region
21882. .22203
/rc_family="L1"
misc_feature
22429. .22797
/note="similar to EST BF993478 (NID:g12399801)"
repeat_region
22717. .23007
/rc_family="L1"
misc_feature
22993. .23191
/note="similar to EST BF767044 (NID:g12115035)"
repeat_region
23167. .23471
/rc_family="Alu"
misc_feature
23511. .23930
/note="similar to EST AW105611 (NID:g6076346) x049g02.x1"
repeat_region
25387. .25677
/rc_family="Alu"
```

## Query Match

28.6% Score 290.6; DB 9; Length 212827;

Best local similarity 64.5% Pred. No. 7e-75;

Matches 664; Conservative 0; Mismatches 304; Indels 61; Gaps 13;

```
OY      1 CCCCTATTCTGCTCAGATTAAAGATTGCCAAATACCTTGTGACATGAAGTTGC---GTT 57
DB      174899 CCCCATCTCTGTGCAGATTAAAGATTGCCAAATACCTTGTGACATGAAGTTGCATTTGTT 174958
OY      58 GTGCTGAGAACACGTAAGCTAAGCTGTTGAGAGACTTTTGTCTTAAGAAGACTGCACG 117
```

Db 174959 GTCCCGAGACACCGGACACTGTAACCTTGCAGAACCTTCCTTTGAGAAAGAGGAGAC 175018  
OY 118 TTCTGGGCTCAGG-GGTCAGACCCGCTCCGCTAGC-----TCCCGAGCCGCTGACACA 169  
Db 175019 TTCTGAGATTGAGAGGTGACGACACTTGCCTCTCTATGTAAGTCTCTAGATGCTGTAAGCA 175078  
OY 170 GCACAGCTCTCTTAATGACAGCGCTGCATGTAAGCAGCTGTAATCACTTATCAGCCATGCT 229  
Db 175079 GAACAGCTCTCCCAATGAAGCGTTGGCATTTGAACCTCAGCAGTAGATGAGACAGCGTT 175138  
OY 230 CATTAAGTACTTCTGTAAGTACGTAACAGAGGCTGTAACAGCTGCTCTCTTGAATTTCA 289  
Db 175139 CCGGACATTAACATTTGACTGTAATGAGAGGCTGAGCAGCTCAGCTCTTTGGATCAGTC 175198  
OY 290 TACTGATCTCTAAATAACACGCTGACGCGCTTGCAGCGCTTGAACGAGACTGCGC 349  
Db 175199 TTGTGATTTCAATGAGAGATTCTTGAACAGCTTTTGGAGATTTGAACAGACTG-- 175256  
OY 350 TCCTGTGTGCTCTAAGCAAGATTTCTGCTAGTCTGAGGCTGCTGAGTGT 409  
Db 175257 ---CATTTCCCTATGAGAAATCTGTT--TACCTGTGGGTGTCGCGGTGGGTGT 175310  
OY 410 GTGAAACAGAGCTCATCAAGAGAGACAGACTATTTGAC--TAAATGAGATAGAGAT 468  
Db 175311 GT-----GTGATCAAGAGCAAGACAGTATTTGACAAATATACAGATGAG-- 175358  
OY 469 TAATTACACTAGTATGATGAGATA--TTCAAGCAATTAAGTGTACGCGTAA 525  
Db 175359 --ATTACACTAGTATGATGAGATAAGTGTGACGCGTAAATCTTAAAGTTA 175416  
OY 526 AGCTTTTAAAGGTTAATTTCTGTCAAGAGTATGACAAATGGCGATCTTATCAGTG 585  
Db 175417 ATTTCTGTCAAAATGAGTATGATGAGAAAGGTTGATATATACGAAATGTTTCT 175476  
OY 586 TCTCTTTAGAGCCCTTCCCTGCTGCTCCCTCCAGATGGGCGTTGAGTCAAT 645  
Db 175477 TAAGCTTTCTCTCTCTTACACCTGCAAGCTCCCAAAATGGGCAATTAATTCATCT 175536  
OY 646 TTAACCTGGCATCTCTACAGTGTGTAACCTTACAGTGTCTTTTACAGTCCCTTC 705  
Db 175537 TTAACCTGGTGTCTCTGTAAGTGTGTAAGTGTCTTTCTTATGAAACCCCTTC 175596  
OY 706 TTAACGAGCAATATGCTGACCTGTAATATGATCTTCTGATAATGCAATTCGAGAT 765  
Db 175597 TGACGAGCAATATGCT--CCTGATATTAATTAATCTTCTGATATGCAATTAAGGCT 175655  
OY 766 TTTTGTGATATGTAAGAGTGTCTGCTTTTACCTTCTTACTACGCTGACTAGT 825  
Db 175656 TTTTGTGATATGTAAGAGTGTCTTCTCAATGTTAC-----TTTATTCAGAGCAATTAAGT 175710  
OY 826 GCTTCCCTGCTTTTCTAGTAAGTGTGTAAGTGTGTAAGTGTGTAAGTGTGTAAGT 885  
Db 175711 GCTTCTCTTACTTTTCTAGTAAGTGTGTAAGTGTGTAAGTGTGTAAGTGT 175770  
OY 886 TAAACTATTTTAAATA-----TCTGAAACATCACTGCTTCTGCAAG 928  
Db 175771 TAAATAATTTTAAATAATCTTAAACTATGAACTCTTAACTACTGCTTCTGCAAG 175830  
OY 929 TAACTACTGCTATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 988  
Db 175831 TAACTACTGCTATGATGATGATGATGATGATGATGATGATGATGATGATGAT 175889  
OY 989 CTTCGGTA 997  
Db 175890 CCTCTGTA 175898

RESULT 15  
AX408985 1232 bp DNA linear part 14-JUN-2002  
LOCUS AX408985  
DEFINITION Sequence 1632 from Patent W00229103.  
ACCESSION AX408985

VERSION AX408985.1 GI:21441690  
KEYWORDS Homo sapiens (human)  
SOURCE Homo sapiens  
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE  
1  
Alvares, C., Horne, D., Peres-da-Silva, S. and Vockley, J.G.  
Gene expression profiles in liver cancer  
Patent: WO 0229103-A 1632 11-Apr-2002;  
GENE LOGIC INC (US)  
FEATURES  
1.1232  
Location/Qualifiers  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
/note="EMBL/Genbank Accession No. D25274"  
BASE COUNT 329 a 240 c 251 g 411 t 1 others  
ORIGIN  
Query Match 28 0%; Score 285.2; DB 6; Length 1232;  
Best Local Similarity 64.4%; Pred. No. 1.9e-73;  
Matches 658; Conservative 0; Mismatches 303; Indels 61; Gaps 13;  
OY 8 TCTTGTCAATTAAGATTTGCCAAATACCTTGTGAAGTATGCT---GTGTGCTGA 64  
Db 1 TCTTGTCAATTAAGATTTGCCAAATACCTTGTGAAGTATGCT---GTGTGCTGA 60  
OY 65 GAACAGCTAAGCACTAAGCTGTTGAGAGACTTTGCTTTAAGAAAGCTGCTGTGG 124  
Db 61 GAACAGCTAAGCACTAAGCTGTTGAGAGACTTTGCTTTGAAAGAGAGCTGTGCA 120  
OY 125 CTCAGG-GGTGAGACCCCTCCGTTAGC-----TCCAGACGCTGAGACAGCAGC 176  
Db 121 GTTAGAGAGGTGACAGCTTCTCTATATGATTTCTGAGTGCAGAAAGAGCAGC 180  
OY 177 CTCTTAATGACAGCTGCTCATGTAAAGCACTGTATATGAGCCATGCTCATAGC 236  
Db 181 CTCCGCAATGAAGCTTGCATTTGAACTGACAGTGTGAGCAGCAGCTGTTCCGACA 240  
OY 237 TAATTTGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 296  
Db 241 TAATTTGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 300  
OY 297 TTTCTTAATTAACAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 356  
Db 301 TTTCTTAATTAACAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 354  
OY 357 TTGCTCTTAAGAGATTTCTGCTAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCT 416  
Db 355 TTTCTCTTAATTAAGAGATTTCTGCTAGTGTGCTGCTGCTGCTGCTGCTGCTGCT 407  
OY 417 ACAGCTGATCAAGAGAGACAGCTATTTTAC--TAATATGAGTATGATTAATTA 475  
Db 408 -----GTATCAAAAGAGCAAGAGCTATTTTACAAATACAAAGTGTGAG--ATTA 458  
OY 476 CACTACTGTATGATGAGT---AATTTCAAGCAATTAAGTGTGCTGCTGCTGCT 532  
Db 459 CACTACTGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 518  
OY 533 TAAAGGTATTTTCTGTAACAGATGATGATGATGATGATGATGATGATGATGATGAT 592  
Db 519 TCAATATGAGTATGATGATGATGATGATGATGATGATGATGATGATGATGAT 578  
OY 593 TGAAGCCCTTCTCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 652  
Db 579 TTTCTTCTCTTACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 638  
OY 653 GGCATCTGACAGTGTGATGATGATGATGATGATGATGATGATGATGATGATGAT 712  
Db 639 GGTGTGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 698  
OY 713 GCAATATGCTGACCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 772

```
Db      699  GCAATATGCTT-CCTGTATATATAAACTTCTGATATAGCATTAGAAGGTTTTTGT 757
QY      773  TAGATAGTAGAAGTGCCTTCTGTTTCACCTTCTTACTCAGCTGACTAGTCTCC 832
Db      758  CGATTAGTAAAGTGCCTTCCATGTAC-----TTATTCAGAGCTAATAGTCTTCC 812
QY      833  TTCGTTTCTAGTACTGGGTGTAGAAATCAGTGTGCGGCTTTACAGTTTAAACTA 892
Db      813  TTAGTTTCTAGTACTAGGTAAATCATGTGTGACGCTTTATAGTTTAAATA 872
QY      893  TTTAGATA-----TTCGAACATCATGCTCTGCGCAGATACCAAC 935
Db      873  TTTAGATTAATTTTAACATATGAACCTTCTTAACATCATGCTGTGCGAGATTACGAC 932
QY      936  ACTGTCAATGATATGATGCGCCCTCTAGACCTCAACCCAGCGGACACATGCTCCGG 995
Db      933  ACTGTCACTTGACCAATACTG-AACCTCTTACCTCGGCCACGGGAGACACGCTCCTG 991
QY      996  TA 997
Db      992  TA 993
```

Search completed: August 26, 2003, 02:30:03  
Job time : 3758.19 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 25, 2003, 19:52:32 ; Search time 280.632 Seconds  
(without alignments)  
9782.674 Million cell updates/sec

Title: US-09-717-321a-15

Perfect score: 1017  
Sequence: 1 ccccatctctgtcagatc.....ccttggtctgtgagttc 1017

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 2552756 seqs, 1349719017 residues

Total number of hits satisfying chosen parameters: 5105512

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : N\_Geneseq\_19Jun03:\*

1: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT:\*  
2: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT:\*  
3: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT:\*  
4: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT:\*  
5: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT:\*  
6: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT:\*  
7: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT:\*  
8: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT:\*  
9: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT:\*  
10: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT:\*  
11: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT:\*  
12: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT:\*  
13: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT:\*  
14: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT:\*  
15: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT:\*  
16: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT:\*  
17: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT:\*  
18: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT:\*  
19: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT:\*  
20: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:\*  
21: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:\*  
22: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:\*  
23: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT:\*  
24: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT:\*  
25: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2003.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	1017	100.0	1017	22	AAH22397	Human rac1 gene re
2	1017	100.0	1266	22	AAH22396	Human rac1 contig
3	290.6	28.6	1022	22	AAH22398	Human rac1 gene re
4	290.6	28.6	2051	21	AAH18236	Human cancer associ
5	290.6	28.6	28567	25	ABH17030	Human MP21 gene Ra
6	285.2	28.0	1232	24	ABK83623	Human CDNA differe
7	285.2	28.0	1232	24	ABH95134	Gene #1632 used to
8	193.2	19.0	605	24	ABH38065	Human colon tumour

C	9	187.2	18.4	605	21	AAH16478
	10	172.2	16.9	432	25	ABX42447
	11	168	16.5	1318	23	AAH65569
	12	164	16.1	2167	24	ABH35389
	13	126.4	12.4	285	24	ABH71132
	14	123.2	12.1	416	25	ABH43616
	15	113	11.1	348	21	AAH22399
	16	111.6	11.0	447	21	AAH89693
	17	111.2	10.9	422	21	AAH89694
	18	103.2	10.1	3740	22	AAH07296
	19	98.2	9.7	581	24	ABH87661
	20	91.6	9.0	565	24	ABH86985
	21	87.8	8.6	506	22	AAH08004
	22	87.6	8.6	470	22	AAH18145
	23	85.6	8.4	466	22	AAH25750
	24	85.6	8.4	495	24	ABH37914
	25	68	6.7	588	21	AAH16128
	26	67	6.6	594	24	ABH60324
	27	62.2	6.1	426	25	ABH42177
	28	61.4	6.0	378	21	AAH06949
	29	59	5.8	404	25	ABH49041
	30	52.4	5.2	162	24	ABH45998
	31	41.4	4.1	901	25	ABH16468
	32	39.2	3.9	4081	23	ABH12618
	33	37.4	3.7	13712	24	ABH33531
	34	37	3.6	352	22	AAH24400
	35	37	3.6	2277	25	ABH58242
	36	37	3.6	260209	24	ABH56564
	37	36.8	3.6	3014	23	ABH06602
	38	36.8	3.6	3724	23	ABH20520
	39	36.6	3.6	995	22	AAH15320
	40	36.4	3.6	7892	24	ABH40055
	41	36	3.5	5098	22	AAH13425
	42	36	3.5	14041	22	AAH48024
	43	36	3.5	16258	24	ABH70376
	44	36	3.5	16258	24	ABH40038
	45	36	3.5	611590	21	AAH22303

#### ALIGNMENTS

RESULT 1  
ID AAH22397 standard; DNA; 1017 BP.  
AC AAH22397;  
XX  
XX 22-AUG-2001 (first entry)  
XX  
XX Human rac1 gene related nucleotide sequence #1.  
DE  
XX  
XX Identification; toxic; hepatotoxic; differential gene expression;  
KW NSAD; non-steroidal antiinflammatory drug; ds.  
XX  
XX Homo sapiens.  
XX  
XX WO200138579-A2.  
XX  
XX 31-MAY-2001.  
XX  
XX 21-NOV-2000; 2000WO-US32049.  
XX  
XX 22-NOV-1999; 99US-0166923.  
XX  
XX 18-FEB-2000; 2000US-0183531.  
XX  
XX 20-NOV-2000; 2000US-0717321.  
XX  
XX (CURA-) CURAGEN CORP.  
XX  
XX Gould-Rothberg BE, DiIppio VA, Ramseh TM, Getweil RW;  
XX  
XX WPI; 2001-355948/37.  
XX

Human colon cancer  
Bovine EST associa  
DNA encoding novel  
Human gene express  
Cott fassel-derlve  
Bovine EST associa  
Human rac1 gene re  
Mouse Excl03 nucle  
Mouse Rab2 nucleot  
Human reproductiv  
Human colon cancer  
Human colon cancer  
Human breast cance  
Human breast cance  
Human breast cance  
Human colon tumour  
Human colon cancer  
Human cancer relat  
Bovine EST associa  
Human secreted pro  
Bovine EST associa  
CDNA encoding colo  
Human intracellula  
Drosophila melanog  
Human immune syste  
Human rac1 gene re  
Human G-protein co  
Human SULF2 genom  
Drosophila melanog  
Human breast cance  
Human breast cance  
Human breast cance  
Murine kit/stem ce  
Internal control B  
Chemically treated  
Human chemically p  
Arabidopsis thalia

PT Screening hepatotoxic agent comprises contacting test cell population  
PT expressing RISKMARKER or INJURYMARKER with agent, comparing expression  
PT with reference population and identifying difference in expression  
PT levels -

XX Disclosure; Page 7-8; 76pp; English.

XX The present invention describes a method of screening a test agent for  
CC hepatotoxicity. The method comprises: (a) providing a test cell  
CC population comprising a cell capable of expressing one or more nucleic  
CC acid sequences selected from the group consisting of RISKMARKER 1-8  
CC and INJURYMARKER 1-10; (b) contacting the test cell population with a  
CC test agent; (c) measuring expression of one or more of the nucleic  
CC acid sequences in the test cell population; (d) comparing the  
CC expression of the nucleic acid sequence in the test cell population to  
CC the expression of the nucleic acid sequence in a reference cell  
CC population comprising at least one cell whose exposure status to a  
CC hepatotoxic agent is known; and (e) identifying a difference in  
CC expression levels of the RISKMARKER or INJURYMARKER sequences, if  
CC present, in the test cell population and reference cell population.  
CC The method is useful for identifying a hepatotoxic agent. The present  
CC sequence is given in the exemplification of the present invention.

SQ Sequence 1017 BP; 245 A; 245 C; 216 G; 311 T; 0 other;

Query Match 100.0%; Score 1017; DB 22; Length 1017;

Best Local Similarity 100.0%; Pred. No. 6.9e-308;

Matches 1017; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 CCCCTATTCTTGGTCAGATTTAAGAAATGGCCAAATACCTGTGAACTAAGTGGCTGTG 60  
DB 1 CCCCTATTCTTGGTCAGATTTAAGAAATGGCCAAATACCTGTGAACTAAGTGGCTGTG 60  
OY 61 CTGAGAACACGTAAGACACTAAGCTGTGAGAGACTTGTCTTAAAGAGACTGACGCTTC 120  
DB 61 CTGAGAACACGTAAGACACTAAGCTGTGAGAGACTTGTCTTAAAGAGACTGACGCTTC 120  
OY 121 TGGGCTGAGGGGTGCAACCCCTCCGTAAGCTCCAGACCGGTGACACAGACAGCCCTCC 180  
DB 121 TGGGCTGAGGGGTGCAACCCCTCCGTAAGCTCCAGACCGGTGACACAGACAGCCCTCC 180  
OY 181 TTAATGACACGCTGCCATGTAAACGACCTGTAACTATACAGCCCATCTATTACGTAAAC 240  
DB 181 TTAATGACACGCTGCCATGTAAACGACCTGTAACTATACAGCCCATCTATTACGTAAAC 240  
OY 241 TTTTACTGTACGTACAGATGGGTGAACAGCTCTGCTTTGATTCATAGTAGTTC 300  
DB 241 TTTTACTGTGTACGTACAGATGGGTGAACAGCTCTGCTTTGATTCATAGTAGTTC 300  
OY 301 CTAAATATACGACGCTGACCGGCTTGTGACAGAGAACTGCGCTCTGCTGTTGC 360  
DB 301 CTAAATATACGACGCTGACCGGCTTGTGACAGAGAACTGCGCTCTGCTGTTGC 360  
OY 361 CTCTAACGAAGTATTTCTTCTAGTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTG 420  
DB 361 CTCTAACGAAGTATTTCTTCTAGTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTG 420  
OY 421 CGTATATAAAGGAGACAGACAGTATTTTACTAATATGAGATGAGATTAATTACACTA 480  
DB 421 CGTATATAAAGGAGACAGACAGTATTTTACTAATATGAGATGAGATTAATTACACTA 480  
OY 481 CATGTGATGAGATTAATTAAGTAAGTAAGTGTACAGGCTTAACCTTTTAAAGCTT 540  
DB 481 CATGTGATGAGATTAATTAAGTAAGTAAGTGTACAGGCTTAACCTTTTAAAGCTT 540  
OY 541 AATTTCTGTCAAAACAGTAGATGACAATGGCCGATCTTATAGTGTCTCTTGTAGCCCC 600  
DB 541 AATTTCTGTCAAAACAGTAGATGACAATGGCCGATCTTATAGTGTCTCTTGTAGCCCC 600  
OY 601 CTTTCCCCCTGCTGCTCCCTCCAGATGGGGCTGTGAGTGCATATTTAACTGGCCATCC 660  
DB 601 CTTTCCCCCTGCTGCTCCCTCCAGATGGGGCTGTGAGTGCATATTTAACTGGCCATCC 660

OY 661 TCACAGTTGCTACTTAAGCAAGTCGTTTCTTTAGAGACCCCTCTTAACGACCAATATG 720  
DB 661 TCACAGTTGCTACTTAAGCAAGTCGTTTCTTTAGAGACCCCTCTTAACGACCAATATG 720  
OY 721 TCTGACCTGTACTTAAGATCTTCTGATATATGATCGGAGATTTTGGTGTGATAGT 780  
DB 721 TCTGACCTGTACTTAAGATCTTCTGATATATGATCGGAGATTTTGGTGTGATAGT 780  
OY 781 AGAAGTCGTTCTGTTTTCACCTTCTTACTGAGTACTAGTGTCCCTCTGTTT 840  
DB 781 AGAAGTCGTTCTGTTTTCACCTTCTTACTGAGTACTAGTGTCCCTCTGTTT 840  
OY 841 CTACTAATCGGGGTATAGAAATACAGTGTGCGGCTTTTACAGTTTAACTATTATGAT 900  
DB 841 CTACTAATCGGGGTATAGAAATACAGTGTGCGGCTTTTAACTATTATGAT 900  
OY 901 ATTGTGAACATCACTGCTCTTCCAGAGTACCAACACTGTCATGTATGATGCGGCC 960  
DB 901 ATTGTGAACATCACTGCTCTTCCAGAGTACCAACACTGTCATGTATGATGCGGCC 960  
OY 961 CTCTAGACCTTACCCAGCGGACACATGCTTCCGTAACCTTTGGGTGTGAGGTTTC 1017  
DB 961 CTCTAGACCTTACCCAGCGGACACATGCTTCCGTAACCTTTGGGTGTGAGGTTTC 1017

RESULT 2

AAH2396/c

ID AAH2396 standard; DNA; 1266 BP.

AAH2396;

22-AUG-2001 (first entry)

Human rac1 conflg SEQ ID NO:2.

Identification; toxic; hepatotoxic; differential gene expression;

NSAID; non-steroidal antiinflammatory drug; ds.

Homo sapiens.

W0200138579-A2.

31-MAY-2001.

21-NOV-2000; 2000WO-US32049.

22-NOV-1999; 99US-0166923.

18-FEB-2000; 2000US-0183531.

20-NOV-2000; 2000US-0717321.

(CURA-) CURAGEN CORP.

Gould-Rothberg BE, D'Alipio VA, Ramsehl TM, Gerweln RW;

WPI; 2001-355948/37.

Screening hepatotoxic agent comprises contacting test cell population

expressing RISKMARKER or INJURYMARKER with agent, comparing expression

levels -

Disclosure; Page 7; 76pp; English.

XX The present invention describes a method of screening a test agent for  
CC hepatotoxicity. The method comprises: (a) providing a test cell  
CC population comprising a cell capable of expressing one or more nucleic  
CC acid sequences selected from the group consisting of RISKMARKER 1-8  
CC and INJURYMARKER 1-10; (b) contacting the test cell population with a  
CC test agent; (c) measuring expression of one or more of the nucleic  
CC acid sequences in the test cell population; (d) comparing the  
CC expression of the nucleic acid sequence in the test cell population to  
CC the expression of the nucleic acid sequence in a reference cell  
CC population comprising at least one cell whose exposure status to a



CC hepatotoxic agent in known; and (e) identifying a difference in  
 CC expression levels of the RISKMARKER or INJURYMARKER sequences, if  
 CC present, in the test cell population and reference cell population.  
 CC The method is useful for identifying a hepatotoxic agent. The present  
 CC sequence is given in the exemplification of the present invention.

XX Sequence 1266 BP; 385 A; 258 C; 285 G; 338 T; 0 other;

Query Match Best Local Similarity 100.0%; Score 1017; DB 22; Length 1266;  
 Matches 1017; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CCCCTATCTCTGCTCAGATTAGAAATGGCAAAATACCTTGAGAACTAGTGGCTGTG 60  
 DB CCCCTATCTCTGCTCAGATTAGAAATGGCAAAATACCTTGAGAACTAGTGGCTGTG 1202  
 QY 61 CTGAGACACAGTAAAGACATAGCTGTGAGAGACTTGTCCCTTAAGAAACATGACCTTC 120  
 DB 1201 CTGAGAACACGTAAAGACATAGCTGTGAGAGACTTGTCCCTTAAGAAACATGACCTTC 1142  
 QY 121 TGGGGCTCAGGGGTGACAGACCTCCCGTACTCCACAGACGCTGAGACACAGACGCTCC 180  
 DB 1141 TGGGGCTCAGGGGTGACAGACCTCCCGTACTCCACAGACGCTGAGACACAGACGCTCC 1082  
 QY 181 TTAATGACACGCTGCCATGTAAAGCAGCACTGTAACTTATCAGCCATCTCATTTACCTAAC 240  
 DB 1081 TTAATGACACGCTGCCATGTAAAGCAGCACTGTAACTTATCAGCCATCTCATTTACCTAAC 1022  
 QY 241 TTTGTACTGTAGCTACAGATGGGTGTAAACAGCTGTCTTTGATTTCAATGAGATTC 300  
 DB 1021 TTTGTACTGTAGCTACAGATGGGTGTAAACAGCTGTCTTTGATTTCAATGAGATTC 962  
 QY 301 CTAAATATACAGCTGACCGGCTTTCGACAGCTTTTGAACAGAACTCGGCTCCGCTGTG 360  
 DB 961 CTAAATATACAGCTGACCGGCTTTCGACAGCTTTTGAACAGAACTCGGCTCCGCTGTG 902  
 QY 361 CTCTAACGAGATTTCTGTCTAGTGTGGGTGTGCTGGGTGAGAGTGTGTAACAGCA 420  
 DB 901 CTCTAACGAGATTTCTGTCTAGTGTGGGTGTGCTGGGTGAGAGTGTGTAACAGCA 842  
 QY 421 GGTCAATCAAGAGACAGACAGTATTTTACATAATGTAAAGTAGAGATTAATTTACACTA 480  
 DB 841 GGTCAATCAAGAGACAGACAGTATTTTACATAATGTAAAGTAGAGATTAATTTACACTA 782  
 QY 481 CATTGTACATGAGATTAATCAACTGAAATGAAGTGCACGGGTAAGCTTTTAAAGGTT 540  
 DB 781 CATTGTACATGAGATTAATCAACTGAAATGAAGTGCACGGGTAAGCTTTTAAAGGTT 722  
 QY 541 AATTTCTGTCAAAAGTAGATGACAAATGGCCGATCTTATCAGTGTCTCTTGAAGCC 600  
 DB 721 AATTTCTGTCAAAAGTAGATGACAAATGGCCGATCTTATCAGTGTCTCTTGAAGCC 662  
 QY 601 CCTTCCCTCTGCTGCTCCCTCCCAAGATGGGGGTGAGTCCATATTTAAACATGCGCATCC 660  
 DB 661 CCTTCCCTCTGCTGCTCCCTCCCAAGATGGGGGTGAGTCCATATTTAAACATGCGCATCC 602  
 QY 661 TCACAGTGTCTAACTTAGCAAGTGTCTTTCTTTAGACACCCCTTTTAAAGCAATATG 720  
 DB 601 TCACAGTGTCTAACTTAGCAAGTGTCTTTCTTTAGACACCCCTTTTAAAGCAATATG 542  
 QY 721 TCTGACCTGTACTATTAAGATCTTTCTGATTAATGCAATTCGAGATTTTGTGTAGATG 780  
 DB 541 TCTGACCTGTACTATTAAGATCTTTCTGATTAATGCAATTCGAGATTTTGTGTAGATG 482  
 QY 781 AAGAGTGGCTCTCTGTTTACACCTTCTTACTACAGCTACATAGTCTCCCTGCTTT 840  
 DB 481 AAGAGTGGCTCTCTGTTTACACCTTCTTACTACAGCTACATAGTCTCCCTGCTTT 422  
 QY 841 CTAGTAACTGGGTGTGAATACAGTGTGCGGCTTTACAGTGTTTTAACTATTTAGAT 900  
 DB 421 CTAGTAACTGGGTGTGAATACAGTGTGCGGCTTTACAGTGTTTTAACTATTTAGAT 362  
 QY 901 ATTCTGAACATCACTGTCTTGCCAGATGACCAACACTGTCAATGTATGTATGCGCCGCC 960

DB 361 ATTCGAAACATCACTGTCTTGCCAGAGTACCAACACTGTCAATGTATGATGCCGCC 302  
 QY 961 CTCTAACCTCAACCCAGGGGACACATGCTCCCGTACCTTTGGCTGTGAGCTTC 1017  
 DB 301 CTCTAACCTCAACCCAGGGGACACATGCTCCCGTACCTTTGGCTGTGAGCTTC 245

# RESULT 3

AAH223398  
 ID AAH223398 standard; DNA; 1022 BP.  
 XX  
 AC AAH223398;  
 XX  
 DT 22-AUG-2001 (first entry)  
 XX  
 XT Human rac1 gene related nucleotide sequence #2.  
 DE  
 XX Identification; toxic; hepatotoxic; differential gene expression;  
 KW NSAID; non-steroidal antiinflammatory drug; ds.  
 XX  
 OS Homo sapiens.  
 XX  
 PN W0200138579-A2.  
 XX  
 PD 31-MAY-2001.  
 XX  
 PF 21-NOV-2000; 2000MO-US32049.  
 XX  
 PR 22-NOV-1999; 990S-0166923.  
 PR 18-FEB-2000; 2000US-0183531.  
 PR 20-NOV-2000; 2000US-0717321.  
 XX  
 PA (CURA-) CURAGEN CORP.  
 XX  
 PI Gould-Rothberg BE, Dipippo VA, Ramseh TM, Gerwein RM;  
 XX WPI: 2001-355948/37.  
 XX  
 PT Screening hepatotoxic agent comprises contacting test cell population  
 PT expressing RISKMARKER or INJURYMARKER with agent, comparing expression  
 PT with reference population and identifying difference in expression  
 PT levels  
 XX  
 PS Disclosure; Page 7-8; 76pp; English.  
 XX  
 CC The present invention describes a method of screening a test agent for  
 CC hepatotoxicity. The method comprises: (a) providing a test cell  
 CC population comprising a cell capable of expressing one or more nucleic  
 CC acid sequences selected from the group consisting of RISKMARKER 1-8  
 CC and INJURYMARKER 1-10; (b) contacting the test cell population with a  
 CC test agent; (c) measuring expression of one or more of the nucleic  
 CC acid sequences in the test cell population; (d) comparing the  
 CC expression of the nucleic acid sequence in the test cell population to  
 CC the expression of the nucleic acid sequence in an reference cell  
 CC population comprising at least one cell whose exposure status to a  
 CC hepatotoxic agent in known; and (e) identifying a difference in  
 CC expression levels of the RISKMARKER or INJURYMARKER sequences, if  
 CC present, in the test cell population and reference cell population.  
 CC The method is useful for identifying a hepatotoxic agent. The present  
 CC sequence is given in the exemplification of the present invention.  
 XX  
 SQ Sequence 1022 BP; 273 A; 208 C; 205 G; 336 T; 0 other;

Query Match Best Local Similarity 28.6%; Score 290.6; DB 22; Length 1022;  
 Matches 664; Conservative 0; Mismatches 304; Indels 61; Gaps 13;

QY 1 CCCCTATCTCTGCTCAGATTAGAAATGGCAAAATACCTTGGAAGTAACTAGTGC---GTT 57  
 DB 1 CCCCTATCTCTGCTCAGATTAGAAATGGCAAAATACCTTGGAAGTAACTAGTGC---GTT 60  
 QY 58 GTGCTGAGAACAGTAAAGCACTGTGAGAGACTTGTCCCTTAAGAAAGACTGACAGC 117

```

Db      61  GTGCGGAGAACACCGAGCACTGAACTTTGCAAAAGACCTTCTTTTGAAGAGCGTAGC 120
QY      118  TTCTGGGCTCAGG-GGTGACAGACCCCTCCGTAAGC-----TCCAGACCGGTGTGACACA 169
Db      121  TTCTGAGCTT AGGAGGTGACAGACACTGTCTCTCTATGTAGTGTCTGAGATGCCGTAAGACA 180
QY      170  GCACAGCCCTCTTAATATACAGCGTGCATATGTAACGCACCTGTAACTTATACGCCCATGCT 229
Db      181  GAACAGCCCTCCGAGATGAAAGCGTTGGCCATGTGAACTCAGAGTGATGAGACAGCGTGT 240
QY      230  CATTACGTAATTTGTACTGTACTGTACGTACAGATGGGTGTAACAGCTCTGCTCTTGTATTCA 289
Db      241  CCGGACATTAACATTTGTCTCTTAATGAGTAGAGCTACAGCTCAGCTCTTTGGATTCAGTC 300
QY      290  TAGTGAATCTCTAATAATACAGAGCTGACCGGCTTCTGACAGGCTTTGAAACAGAACTCTGC 349
Db      301  TTGTGATTTCAATGAGCAGATTTTCTGACAGCAAGCTTTTGGGAGATTTTGAACAGAACTG 358
QY      350  TCCGTGTGTGCTCTTAACAGAGTATTTCTGTCTCTAGTCGTGGGTGTCTGTGGGTGAGTGT 409
Db      359  ---CTATTCTCTTAATGAAATTTCTGTT--TAGCTGTGGGTGTCCGGGTGGGGTGT 412
QY      410  GTGAACACGACGTCATCAAAAGAGACAGACAGATTTTGTGAC-TAATATGAAAGTAGAGAT 468
Db      413  GT-----GTGATCAAAAGACAAAGACAGATTTTGTACAAATATACAGAGTGGAG-- 460
QY      469  TAATTACACTATCTTATGATGAGTAA--TTCAACTGAAATAAAGTGTACAGGTTAA 525
Db      461  -ATTATACATACATTTGTACAGGAATGAAAGTGTACAGGTTAAAGTCTTAAAGAGTTA 518
QY      526  AGCTTTTAAACGTTAATTTCTGTCAAAACAGTAGATACAAATGGCCGACTTTTCAAGTG 585
Db      519  ATTCTCTCAATGAGCATAGATGATGTAAGAAAGGTGGTGTATTCAGAAATTTTCT 578
QY      586  TCTCTCTTGAAGCCCTCCCTCCCTGCTGTCCTCCAGATGAGGCGGTTGAGTCCATAT 645
Db      579  TAAGCTTTCTCTCTTACACTGCGATGCGTCCGCCAAATTTGGCATTAATTCATCT 638
QY      646  TTAACATGGCCATCTCTACAGTGTCTAATTAAGTAAGTCTTTTCTTTAGAGACCCCTTC 705
Db      639  TTAACATGGTGTCTCTGTAGTCCCTAATAGTAAGTGTCTTTCTTATAGAACCCCTTC 698
QY      706  TTAACGAGCAATATGTGACCTGTACTATTAAGATCTTTTGATTAATGATCTGGAGATT 765
Db      699  TGACTGAGCAATATGCTCT-CTTGTATTAATTAATCTTTTGATTAATGATTAAGAGTT 757
QY      766  TTTTGTGATAGTATGAGAGTGCCTCTCTGTTTTCACCTTCTCTTACTACAGCTGACTAGT 825
Db      758  TTTTGTGATAGTATGAGAGTGCCTCTCTGTTTTCACCTTCTCTTACTACAGCTGACTAGT 812
QY      826  GCTTCCCTCTGTTTCTTGTAGTACTGAGTGGTGTGAAATACAGTGTGCGGCTTTACAGTTT 885
Db      813  GCTTCCCTGTTTCTTGTAGTACTGAGTGGTGTGAAATACAGTGTGCGGCTTTATAGTTT 872
QY      886  TAAACATTTTGTAGTA-----TTGGAACATGACACTGTCTTCCAGAG 928
Db      873  TAAATATTTTGTAGTAAATTTCTTAACATATGAACTTCTTAACATACACTGTCTTCCAGAT 932
QY      929  TACCAACACTGTCTGTATGATGCCGCCCTCTAGACCTCAGCCACAGCGGAGACAGATG 988
Db      933  TACGACACGTCTGACTGTGACCAATATGAG-ACCCTCTTACCTGCGCCACGCGGAGACAGC 991
QY      989  CTTCCGGTA 997
Db      992  CTTCTCTGTA 1000

```

RESULT 4  
 AAF18236  
 ID AAF18236 standard; DNA; 2051 BP.  
 XX  
 AC AAF18236;

```

XX      14-MAR-2001 (first entry)
DT
XX
XX      Lung cancer associated polynucleotide sequence SEQ ID 255.
DE
XX
XX      Human; lung cancer associated protein; neuroprotective; cytoskeletal;
KW      cardioactive; immunomodulatory; muscular active; vulnery;
KW      gastrointestinal; nephrotoxic; antineoplastic; gynecological;
KW      antibacterial; diagnosis; neural disorder; immune disorder; reproductive;
KW      proliferative disorder; wound healing; infectious disease; ds.
XX
XX      Homo sapiens.
OS
XX
XX      WO20005180-A2.
FN
XX
XX      21-SEP-2000.
PD
XX
XX      08-MAR-2000; 2000WO-US05918.
PE
XX
XX      12-MAR-1999; 99US-0124270.
PR
XX
XX      (HUMA-) HUMAN GENOME SCI INC.
PA      (ROSE/) ROSEN C A.
PI
XX
XX      Ruben SM.
PI
XX
XX      WPI: 2000-587514/55.
DR
XX
XX      P-PSDB: AAB58360.
PT
XX
XX      Lung cancer associated gene sequences, referred to as lung cancer
PT      antigens, useful for treatment, prevention, and diagnosis of disorders
PT      such as lung cancer -
XX
XX
XX      Claim 1; Page 716-717; 1425pp; English.
PS
XX
XX      Polynucleotide sequences AAF17982 - AAF18424 encode human lung cancer
CC      associated proteins represented in AAB58106 - AAB58548. Lung cancer
CC      associated proteins and polynucleotide sequences, their agonists, and
CC      antagonists may have neuroprotective, cytoskeletal, cardioactive;
CC      immunomodulatory; muscular active general; vulnery; gastrointestinal
CC      general; nephrotoxic; antineoplastic; gynecological; or antibacterial
CC      activity. The invention also includes antibodies specific for the
CC      protein or polynucleotide sequences. The lung cancer associated
CC      polynucleotide sequences may be used for detection of lung cancer,
CC      chromosome identification, as chromosome markers, and for numerous other
CC      diagnostic or research purposes. The proteins may be used to treat
CC      disorders such as neural, immune, muscular, reproductive,
CC      gastrointestinal, pulmonary, cardiovascular, renal, and proliferative
CC      disorders. The proteins may also be used in the treatment of wounds and
CC      infectious diseases. Polynucleotide sequences AAF18425 - AAF18433 and
CC      peptide AAB58549 are used in the course of the invention for the
CC      identification and characterisation of the polynucleotide and protein
CC      sequences.
XX
XX      Sequence 2051 BP; 570 A; 430 C; 433 G; 612 T; 6 other;
SQ

```

Query Match 28.6%; Score 290.6; DB 21; Length 2051;  
 Best Local Similarity 64.5%; Pred. No. 4.5e-80;  
 Matches 664; Conservative 0; Mismatches 304; Indels 61; Gaps 13;

```

QY      1  CCCCTATTCTGTGCATGATTAAGATGTCGCAAAATACCTGTGACATTAAGTGC---GTT 57
Db      755  CCCCATTCTGTGTCATTAATTAAGATGTCGCAAAATACCTGTGACATTAAGTGTGTT 814
QY      58  GTGCTGAGAACACGTAAGCACTAGTGTGAGAGACTTGTCTCTTAAGAGAGTGCAGC 117
Db      815  GTGCGGAGAACACGAGACACTGAACCTTTGCAAAACCTTGTCTTTAGAGAGAGTGCAGC 874
QY      118  TTCTGGGCTCAGG-GGTGACAGACCCCTCCGTAGC-----TCCAGACCGGTGTGACACA 169
Db      875  TTCTGAGATTGAGAGGTGACAGACACTGTCTCTATAGTATGTTCTCAGATGCCGTAAGACA 934
QY      170  GCACAGCCCTCTTAATATGACAGCGTGCATGTAAGCAACCTGTAACTTATACGCCCATGCT 229

```

[illegible]

OS Homo sapiens.  
XX  
XX  
XX WO2003006990-A1.  
XX  
XX  
XX 23-JAN-2003.  
XX  
XX  
XX 10-JUL-2002; 2002WO-US21549.  
XX  
XX 12-JUL-2001; 2001US-305017P.  
XX 10-OCT-2001; 2001US-328491P.  
XX 15-FEB-2002; 2002US-357452P.  
XX  
XX  
XX (EXCEL-) EXELIXIS INC.  
XX  
XX  
XX Friedman L, Plowman GD, Belvin M, Li D, Funke RP;  
XX  
XX WPI; 2003-221779/21.  
XX  
XX P-PSDB; ABJ19756.  
XX  
XX  
XX Identifying candidate p21 pathway modulator, by contacting an assay  
XX  
XX system having modifiers of p21 polypeptide or gene with a test agent to  
XX  
XX provide a reference activity in system and detecting test agent-biased  
XX  
XX activity -  
XX  
XX Examples; Page 56-72; 199pp; English.

[illegible]

```

Db 27229 ----CTATTCTCTTAATGAAGAAATTCGTT--TAGCTGGGGGTGGCGGGTGGGGT 27282
Qy 410 GTGAACACGACGATCATCAAGAGACAGACAGATATTTGAC-TAATATGAGTAGAGAT 468
Db 27283 GT-----GTGATCAAAAGGACAAAGACAGATTTTGCACAAATATGAGAGTGAGC-- 27330
Qy 469 TAAATTAACATACATTTGACATGAGATPAA--TTCACTGAAATAAAGTGTACAGGGTAA 525
Db 27331 --ATTGACATACATTTGACAAAGAAATGAAAGTGTACAGGGTAAATTAAGGTTA 27388
Qy 526 AGCTTTTAAAGGTTAATTTCTGTGCAAAACATAGATGCAATAGCCGATCTTATCACTG 585
Db 27389 ATTCTGTCAATGCAATGATGATGATGAAAGAAAGGTGGTATTTATCAGGAATGTTTCT 27448
Qy 586 TCTCTCTGAGCCCCCTTCCCTGCTGCTGCCAGATGGGGGCTTGAGTCATAT 645
Db 27449 TAAGCTTTTCTTCTTCTTACACCTGCCATGCCCTCCCAATTTGGGCAATTAATTCATCT 27508
Qy 646 TTAACATGCGCATCCTCAGACATGCTTAACCTTAAGCAAGTCTTTCTTTAGACCCCTTC 705
Db 27509 TTAACATGCTGTGTTCTGTGATGCTTAAGCTTAAGTCTTTCTTATAGAACCCCTTC 27568
Qy 706 TTAACGAGCAATATGCTGACCTGACTATTAAGATCTTTGATTAATGCAATTCGAGAT 765
Db 27569 TGACTGAGCAATATGCTCCTTGTATATTAATCTTTGATTAATGCAATTCGAGAT 27627
Qy 766 TTTTGTGATAGTATGAGATGCTGCTGTTTCAACCTCTTACTACAGCTGATAGT 825
Db 27628 TTTTGTGATAGTATTAAGAGTCTTCCATGTTAC-----TTTATTCAGACTAATAGT 27682
Qy 826 GCTTCCCTGCTGTTTCTGATTAACGCTGCTGATGAATACAGTCCGCGCTTACAGTTT 885
Db 27683 GCTTCCCTGCTGTTTCTGATTAACGCTGATGAATACAGTCCGCGCTTACAGTTT 27742
Qy 886 TAACTATTTAGATA-----TTCTGAACATCACTGCTTCCGACAG 928
Db 27743 TAAATATTTAGATAATTTCTTAACATGTAACCTTCTTAATCATCAGCTTCCGACAT 27802
Qy 929 TACCAACATCTGATGATGATGATGCGCCCTCTAGACTACACCGGAGACATG 988
Db 27803 TACCGACATCTGATGATGATGATGCGCCCTCTAGACTACACCGGAGACATG 27861
Qy 989 CTTCGGGTA 997
Db 27862 CTTCCTGTA 27870

```

RESULT 6  
ABR83623  
ID ABR83623 standard; cDNA: 1232 BP.

AC ABR83623;

DE 14-AUG-2002 (first entry)

DE Human cDNA differentially expressed in granulocytic cells #194.

Human; ss; granulocytic cell; DNA chip; bacterial infection;  
viral infection; parasitic infection; protozoal infection;  
fungal infection; sterile inflammatory disease; psoriasis;  
rheumatoid arthritis; glomerulonephritis; asthma; thrombosis;  
cardiac reperfusion injury; renal reperfusion injury; ARDS;  
adult respiratory distress syndrome; inflammatory bowel disease;  
Crohn's disease; ulcerative colitis; periodontal disease;  
granulocyte activation; chronic inflammation; allergy.

OS Homo sapiens.

PN WO200228999-A2.

PD 11-APR-2002.

XX

```

PF 03-OCT-2001; 2001WO-US30821.
XX
PR 03-OCT-2000; 2000US-237189P.
XX
PA (GENE-) GENE LOGIC INC.
XX
PI Beazer-Barclay Y, Weissman SM, Yamaga S, Vockley J;
XX WPI; 2002-435328/46.
XX
PT Detecting granulocyte activation by detecting differential expression
PT of genes associated with granulocyte activation, which serves as
PT diagnostic markers that is useful for monitoring disease states and
PT drug toxicity.
XX
PS Claim 1; SEQ ID No 194; 114pp; English.
XX
CC The invention relates to detecting (M1) granulocyte (GC) activation
CC (GCA), by detecting the level of expression of gene(s) (Gs) identified by
CC DNA chip analysis as given in the specification, and comparing
CC the expression level to an expression level in an unactivated
CC GC, where differential expression of Gs is indicative of GCA.
CC Also included are modulating (M2) GA by contacting GC with an agent
CC that alters the expression of at least one gene in Gs; (2) screening (M3)
CC for an agent capable of modulating GCA or an inflammation (especially
CC chronic) in a tissue, an allergic response in a subject, exposure of a
CC subject to a pathogen or sterile inflammatory disease using the
CC gene expression profile; (3) detecting (M4) an inflammation (especially
CC chronic) in a tissue, an allergic response in a subject, exposure of a
CC subject to a pathogen or sterile inflammatory disease, by detecting the
CC level of expression of the gene(s) of the tissue of Gs, where
CC the level of expression of the gene is indicative of inflammation;
CC (4) treating (M5) an inflammation (especially chronic) or in a tissue,
CC an allergic response in a subject, exposure of a subject to a pathogen
CC or sterile inflammatory disease, by contacting a tissue having
CC inflammation with an agent that modulates the expression of gene(s)
CC from Gs in the tissue. M1 is useful for detecting GCA; M2 is useful for
CC modulating Gs; M3 is useful for screening an agent capable of modulating
CC GCA preferably in an inflammation in a tissue; M4 is useful for
CC detecting an inflammation (especially chronic) in a tissue, an allergic
CC response in a subject, exposure of a subject to a pathogen or sterile
CC inflammatory disease (e.g. psoriasis, rheumatoid arthritis,
CC glomerulonephritis, asthma, thrombosis, cardiac reperfusion injury, renal
CC reperfusion injury, ARDS, adult respiratory distress syndrome,
CC inflammatory bowel disease, Crohn's disease, ulcerative colitis,
CC periodontal disease; also bacterial infection, viral infection,
CC parasitic infection, protozoal infection, fungal infection and M5 is
CC useful for treating one of the above conditions. The present
CC sequence represents a gene differentially expressed in granulocytes.
CC Note: The sequence data for this patent did not form part
CC of the printed specification, but was obtained in electronic
CC format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 1232 BP; 329 A; 240 C; 251 G; 411 T; 1 other:

```

Query Match 28.0%; Score 285.2; DB 24; Length 1232;  
Best Local Similarity 64.4%; Pred. No. 1.7e-78;  
Matches 658; Conservative 0; Mismatches 303; Indels 61; Gaps 13;

```

Qy 8 TCTTGCTAGATTAAAGTGGCAAAATACCTGTGACATGAAGTTGC--GTGTGCTGCA 64
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1 TCTTGTAGATTAAAGTGGCAAAATACCTGTGACATGAAGTTGTGGCGCA 60
Qy 65 GAACACGTAGACATTAAGCTGTGAGAGACTTGTCTTAAGAAAGACTGACCTTGGG 124
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 61 GAACACGTAGACATTAAGCTGTGAGAGACTTGTCTTAAGAAAGACTGACCTTGGCA 120
Qy 125 CTCAGG-GGTGACAGCCTCCGTAAGC-----TCCAGACGCTGTGACACAGC 176
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 121 GTTAGAGAGTGCAGACACTTGTCTCTATGTAGTTCAGATGCGTAAAGACAGAGC 180
Qy 177 CTCCTTAATGACAGCCTCCATGTAAAGCAGCAGCTTAATTAAGCCCATGCTCATAGC 236

```





KW detect mutation: non-invasive diagnostic method; ds.  
 XX Homo sapiens.  
 OS  
 XX MO200012702-A2.  
 PN  
 XX 09-MAR-2000.  
 PD  
 XX 30-AUG-1999; 99WO-US19424.  
 PF  
 XX 31-AUG-1998; 98US-0098639.  
 PR 27-JAN-1999; 99US-0117393.  
 XX  
 XX (FARB ) BAYER CORP.  
 PA  
 PI Endege WO, Steinhann KE, Astle JH, Burgess CC, Carroll E;  
 PI Catino TJ, Dwyer P, Ford DM, Lewis ME, Molino GA, Monahan JE;  
 PI Schlegel R;  
 XX  
 XX WPI: 2000-256641/22.  
 DR  
 XX  
 XX Novel nucleic acids and proteins for identifying therapeutic agents  
 PT useful for treating and diagnosing cancer, especially colon cancer -  
 XX  
 XX Claim 16; Page 319; 345pp; English.  
 PS  
 XX  
 XX This sequence represents a human nucleotide sequence which is  
 CC differentially expressed in colon cancer cells compared to the expression  
 CC levels in normal cells. The nucleotide sequence can be used as a source  
 CC of primers and probes. The nucleotide sequence is useful for determining  
 CC the phenotype of a cell by detecting the differential expression of the  
 CC sequence relative to a normal cell. The probes derived from the sequence  
 CC can also be used to determine the phenotype of cells in a sample. Probes  
 CC and antibodies which hybridize to the nucleotide sequence can also be  
 CC used to determine the phenotype of a cell. The primers are useful for  
 CC detecting a mutation in a test nucleotide sequence and also for detecting  
 CC cancer, preferably colon cancer. Antibodies against the protein encoded  
 CC by the nucleotide sequence can also be used in a method to detect colon  
 CC cancer. The diagnostic method is non-invasive and accurate for diagnosing  
 CC colon cancer at an early stage.  
 CC  
 XX  
 SQ Sequence 605 BP; 214 A; 105 C; 116 G; 155 T; 15 other:  
 Query Match 18.4%; Score 187.2; DB 21; Length 605;  
 Best Local Similarity 65.5%; Pred. No. 5.9e-48;  
 Matches 348; Conservative 0; Mismatches 158; Indels 25; Gaps 5;  
 OY 485 GTACATGAGATATTCACCTGATTAAGTGCAGGGTAAAGCTTTTAAAGGTTAATT 544  
 DB 597 GGACACANGANNNAAGTGCNGGGTAAAGCTTTAAAGGTTAATTTCGCAATNCAGTA 538  
 OY 545 TCGTCACACAGAG-ATGACAAATGGCGGATCTTATCAGTGTCTCTGAGACCCCT 603  
 DB 537 GATTAANNAANNAAGGTTTNTATACACAGNNAATGTTTCTTANCGTTTCTTNTTCT 478  
 OY 604 TCCCTCTGCTGCCCTCCAGATGGGGGCTTGAGTCATATTAACTGGCATCTCA 663  
 DB 477 AACCTGCGCATGCTCCCAATGGGATTTAATTCATCTTTAAACAGCTTGTCTCT 418  
 OY 664 CAGTGTCTAAGTACAGTCTTCTTTAGGACCCCTCTTAAAGCATATATGCT 723  
 DB 417 TAGTGTCTAAGTACAGTCTTCTTTAGGACCCCTCTGACTGAGCAATATGCT 358  
 OY 724 GACCTGTACTATAGATCTTCTGANTAAATGCAATGCGAATTTTGTGATAGTATA 783  
 DB 357 -CCTGTATATTAATTAATCTTCTGATTAATGCAATTAAGGTTTGTGATATTA 299  
 OY 784 AGTGTCTCTGTTTCACTTCTTCTTACTAGCTAGTACTAGTCTCTCTGTTCTTA 843  
 DB 298 AGTGTCTCTGATGTC-----TTTATAGAGCTAATAAGTCTTCTCTTACTTCTTA 244  
 OY 844 GTAACTGGGTGTGAAGAAATACGCTGCGGCTTTACAGTTTAACTATTTAGATA-- 901  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

DB 243 GTAACTAGGTGTAAAAATCATGTGTGACGCTTAAAGTTTTTAAATATTTAGATAAT 184  
 OY 902 -----TTTGAACATCAGTGTCTTCCAGAGATCCACATGTCATG 946  
 DB 183 TCTTAACATTAAGACCTTCTTAAATCATGCTTCTCCAGATTAACCGACATGCTCACTG 124  
 OY 947 ATTAGCGCGCCCTCTAGACCTCAACGCGGAGACATATGCTCCGTA 997  
 DB 123 ACCAATATCTG-ACCCTTTACTCTGCCCGGACGCGGACACACGCTCTCTGTA 74  
 RESULT 10  
 ID ABX42447  
 AC ABX42447 standard; cDNA; 432 BP.  
 XX  
 XX ABX42447;  
 XX  
 XX 20-FEB-2003 (first entry)  
 DE  
 XX Bovine EST associated with lactation/muscle/fat deposition #7612.  
 DE  
 XX  
 KW Bovine; ss; EST; expressed sequence tag; lactation; LMPD;  
 KW muscle deposition; fat deposition; genome mapping; gene identification;  
 KW gene analysis; cattle breeding.  
 XX  
 XX Bos Taurus.  
 OS  
 XX  
 XX US2002137139-A1.  
 PN  
 XX 26-SEP-2002.  
 PD  
 XX 24-SEP-2001; 2001US-0960352.  
 PF  
 XX 12-JAN-1999; 99US-115707P.  
 PR 11-JAN-2000; 2000US-0480902.  
 XX  
 XX (BYAT/) BYATT J C.  
 PA (MATH/) MATHIALAGAN N.  
 PA (TAON/) TAO N.  
 PA (WAR/) WARREN W C.  
 XX  
 XX Byatt JC, Mathialagan N, Tao N, Warren WC;  
 PI  
 DR WPI: 2003-110599/10.  
 XX  
 PT New nucleic acid associated with lactation, and muscle and fat  
 PT deposition, useful for genome mapping, gene identification and  
 PT analysis, cattle breeding, or for genetically improving cattle -  
 XX  
 PS Claim 2: SEQ ID No 7612; 245pp; English.  
 XX  
 CC The invention relates to a purified nucleic acid molecule associated with  
 CC lactation or muscle and fat deposition (designated LMPD), derived  
 CC from cattle, and the LMPD nucleic acid can specifically hybridize to a  
 CC second nucleic acid molecule comprising any of 15112 nucleotide  
 CC sequences, appearing as ABX34836-ABX49947, or complements of them.  
 CC Also included are: (1) a transformed cell having a nucleic acid  
 CC comprising an LMPD nucleic acid linked to a promoter and a 3' non-  
 CC translated sequence that functions in the cell to cause termination of  
 CC transcription and addition of polyadenylated ribonucleotides to a 3' end  
 CC of the mRNA molecule; and (2) determining a level or pattern of a  
 CC molecule in a bovine cell or tissue comprising: (a) incubating a marker  
 CC nucleic acid (comprising any of the 15112 nucleic acid sequences or its  
 CC complement or fragment) with a complementary nucleic acid molecule  
 CC obtained from the bovine cell or tissue, where hybridisation between the  
 CC marker nucleic acid and the complementary nucleic acid permits the  
 CC detection of the molecule; and (b) detecting the level or pattern of the  
 CC complementary nucleic acid, where the detection of the complementary  
 CC nucleic acid is predictive of the level or pattern of the molecule.  
 CC The LMPD nucleic acid is used for determining a level or pattern  
 CC of a molecule in a bovine cell or tissue. It is useful for genome  
 CC mapping, gene identification and analysis, cattle breeding, preparation  
 CC of constructs for use in cattle gene expression, or for genetically



CC Improving cattle. The present sequence is one of the 15112 bovine  
CC LMD EST (expressed sequence tag) nucleic acids.  
CC Note: The present sequence was not shown in the specification but  
CC was obtained in electronic format from the USPTO web site:  
CC seqdata.uspto.gov/sequence.html?docid=20020137139.  
XX  
SQ Sequence 432 BP; 123 A; 76 C; 78 G; 155 T; 0 other;  
Query Match 16.9%; Score 172.2; DB 25; Length 432;  
Best Local Similarity 72.2%; Pred. No. 2,4e-43;  
Matches 294; Conservative 0; Mismatches 103; Indels 10; Gaps 5;  
OY 486 TACATGAGATATCACTGATTAAGTGCACGGGTAAAGCTTTTAAAGGTTAATT 545  
DB 27 TACACGTAAACATTAATTAATTAAGCATACGGGTAAAGTAAAGTTAAATTT 86  
OY 546 CTGCAAA--CAGTAGATGACAAATGGCCGATCTTATCAGTGTCTCTTGAAGCC-- 600  
DB 87 CTGCAATGCGGTATGAGAAAGGCGGTATATCAGTGTCTTAAAGCTTT 146  
OY 601 -CCTTCCCTGCTGCTCCCTCCAGATGGGCGGTGAGTCCATATTTAACTGGCCATC 659  
DB 147 TCCCTAAACCTGCCACCCCTTCTGAAATTTGGCATTTAATCATCTGACCTGTATC 206  
OY 660 CTCACAGTGTGCTACTAGCAAGTGTCTTTAGGACCCCTTAAAGCAATAT 719  
DB 207 CTCATATGTCTACTAGTATGATGATGCTTTCTTATGACATCTTAAAGCAATAT 266  
OY 720 GTCGACCTGCTACTAGATATGATCTTGTGATATGATTCGAGATTTTGTGATAGATG 779  
DB 267 GCCT-CCTTGTATTAATAATCTTGTGATATGATTAAGAAATTTTGTGATAGATG 325  
OY 780 TAGAAGGCGTCTGCTGCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 837  
DB 326 TAAAGGCGATCTGCTGCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 385  
OY 838 TTTCTAGTAACT-GGCTGATGAATCAAGTCTCGGCTTACAGTT 883  
DB 386 TTTCTAGTAACTAGTATTAATAATCAAGTCTTCTGACCTTATAGTT 432  
RESULT 11  
AAS65569  
ID AAS65569 standard; cDNA: 1318 BP.  
XX  
AC AAS65569;  
XX  
DT 13-FEB-2002 (first entry)  
XX  
DE DNA encoding novel human diagnostic protein #1373.  
XX  
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;  
KW food supplement; medical imaging; diagnostic; genetic disorder; ss.  
XX  
OS Homo sapiens.  
XX  
PN WO200175067-A2.  
XX  
PD 11-OCT-2001.  
XX  
PF 30-MAR-2001; 2001WO-US08631.  
XX  
PR 31-MAR-2000; 2000US-0540217.  
PR 23-AUG-2000; 2000US-0649167.  
XX  
PA (HXSE-) HXSEQ INC.  
XX  
PI Drmanac RT, Liu C, Tang YT;  
XX  
DR WPI: 2001-639362/73.  
XX  
PT P-PSDB; ABG01382.  
New isolated polynucleotide and encoded polypeptides, useful in

PT diagnostics, forensics, gene mapping, identification of mutations  
PT responsible for genetic disorders or other traits and to assess  
PT biodiversity -  
XX  
PS Claim 1; SEQ ID No 1373; 103pp; English.  
XX  
CC The invention relates to isolated polynucleotide (I) and  
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,  
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
CC and gene mapping, and in recombinant production of (II). The  
CC polynucleotides are also used in diagnostics as expressed sequence tags  
CC for identifying expressed genes. (I) is useful in gene therapy techniques  
CC to restore normal activity of (II) or to treat disease states involving  
CC (II). (II) is useful for generating antibodies against it, detecting or  
CC quantitating a polypeptide in tissue, as molecular weight markers and as  
CC a food supplement. (II) and its binding partners are useful in medical  
CC imaging of sites expressing (II). (I) and (II) are useful for treating  
CC disorders involving aberrant protein expression or biological activity.  
CC The polypeptide and polynucleotide sequences have applications in  
CC diagnostics, forensics, gene mapping, identification of mutations  
CC and to produce other types of data and products dependent on DNA and  
CC amino acid sequences. AAS64197-AAS94564 represent novel human  
CC diagnostic coding sequences of the invention.  
CC Note: The sequence data for this patent did not appear in the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
SQ Sequence 1318 BP; 346 A; 268 C; 279 G; 424 T; 1 other;  
Query Match 16.5%; Score 168; DB 23; Length 1318;  
Best Local Similarity 58.1%; Pred. No. 9,4e-42;  
Matches 615; Conservative 0; Mismatches 365; Indels 78; Gaps 15;  
OY 1 CCCCTATCTCTGCAGATTAAGATTCGCAAAATACCTGTGATAGTTGC---GTT 57  
DB 39 CCCCATCTCTGTCAGATTAAGATTCGCAAAATACCTGTGATAGTTGCATGATGTT 98  
OY 58 GTGCTGAGACAGCTAAGCACTAAGCTGTGAGAGCTTGTG-C-CTTAAGAGACTGCAG 116  
DB 99 GTGCGAGAACACCGAGCAGCTGAACCTTGTGCAAAACCTTGTGAGAGAGCGTAG 158  
OY 117 CTTCTGGGCTCAGG-GGTGCGAGACCTCCGTAGC-----TCCAGACCGGTGACA 167  
DB 159 CTTCTGCACTTGAAGAGGTGCGAGACCTTGTCTCTATGATGTTCTCAGATGCGTAAAG 218  
OY 168 CAGCAGACCTCTTAATGACACGCTGCGATGATGACGACCTCTTAATTA-TCAGCCCAT 226  
DB 219 CAGAAAGACCTCCCGAATGAAGCGTGGCATTTGAATCAGCAGATTAAGCAGCAGT 278  
OY 227 GCTCATTAAGTAACTTTGTACTGATGATGATGATGATGATGATGATGATGATGATGAT 286  
DB 279 GTTCCCGCATGAATGTTCTGATGAATGA--GTGAGGTGAGCAGCTCAGCTCTTGA 336  
OY 287 TCATAGTGTCTCTTAATTAATGACAGCTGACGCGCTTGTGACAGTCTTGAACAGACTCT 346  
DB 337 TCAGTCTTTGATTTCTATGACGATTTTCTGACAGCTTTTGGCGGAGATTTGAACAG 396  
OY 347 GGTCTGTGTGCTCTTAACGAGATTTCTGTTCTAGTGTGCTGTGCTGTGCTGTGAG 406  
DB 397 AACTGGCTATTTCTCTTAATGAAGATTTGTGTTAGCTGTGGGTGTGCGGTGGG 456  
OY 407 TGTGTGAAGACGAGCTCATCAAGAGACAGCAGATTTTGTGCTAATTAAGATGAG 466  
DB 457 GGTGTGTGTGTG---ATCAAAAGGACAAAGAGATTAATTT---TGAACAAATATCGG 508  
OY 467 ATTATTTACACTACATTTGATGAGATTAATTAATTAATTAATTAATTAATTAATTAAT 526  
DB 509 AGTGGAGATTTTACCACTACATTTGATGAGAGATTAATTAATTAATTAATTAATTAAT 568  
OY 527 GCTTTTAAAGCTTAATTTCTGTCAAAAGATGATGATGATGATGATGATGATGATGATGAT 586  
DB 569 CTCTAAAGGTTAATTTCTGTGCAATGATGATGATGATGATGATGATGATGATGATGAT 628

QY 587 -----CTCTCTGAGCCGCCCTTCCCTGCTGCTCCCTCCCA 624  
DB 629 TATCAAGGAAGTTTCTTAAAGCCTTCTTCTTACACCCCTGCTCCCTCC 688  
QY 625 GA---TGGGCGGTGAGTCATATTAACAGCCATCCAGAGTTGCTAACTAGCA 681  
DB 689 AATATGGGCGATTAATATGCTTTAAACGTGTTCTTGTAGCGCTTAACCTAGTA 748  
QY 682 GTG---CTTCTTTAGACCCCTTCTTAAAGCAATATGTGACTGTACTATTA 736  
DB 749 AAGTGCCTTCCCTTAAAGACCCCTTCTGACTGAGCAATATGCTT-CTTGTATTA 807  
QY 737 AATCTTCTTGATATGCAATGGAGATTTTGTGTAGATAGTAAAGTGCCTCTGT 796  
DB 808 AATCTTCTTGATATGCAATGGAGATTTTGTGTAGATAGTAAAGTGCCTCTGT 867  
QY 797 TTTACCTCTCTTACTGACGCTAGTGTGCTTCCCTGCTTCTAGTAAAGTGCCT 856  
DB 868 TTTACTT---TATTCAGAGCTTAATAGTCTTCTTGTAGTCTTCTAGTAAAGT 922  
QY 857 GAAATCAGCTGTGCGGCTTTACAGTTTAAACATATTTAGATA----- 901  
DB 923 AAAATCATGTGTGAGGCTTTATAGTTTAAATTTTAAATTTCTTAACATATGA 982  
QY 902 --TTCTGAAACATCACTGTCTTGTCCAGAGTACACACTGTCTATGATGATGCC 959  
DB 983 ACCCTTCTTACATCACTGTCTTGTCCAGATTCGACACACTGTCTGACCAATATG 1041  
QY 960 CCTTAGACCTCACCACGCGGACACATGCTTCCGGTA 997  
DB 1042 CCTCTTACCTGCGCCACGCGGACACACCTCTCTGTA 1079

RESULT 12  
AB235389  
ID AB235389 standard; cDNA; 2167 BP.  
XX  
AC AB235389;  
XX  
DT 05-FEB-2003 (first entry)  
XX  
DE Human gene expression profile polynucleotide SEQ ID NO 500.  
XX  
KW Human; artery; endothelium; umbilical; vein; aorta; pulmonary artery;  
KW bronchial epithelium; prostate; muscle; lung fibroblast; osteoblast;  
KW tumour; microarray; genome mapping; antibiotic; antiviral; antifungal;  
KW gene expression; gene; ss.  
XX  
OS Homo sapiens.  
XX  
PN WO200274979-A2.  
XX  
PD 26-SEP-2002.  
XX  
PF 20-MAR-2002; 2002WO-US08456.  
XX  
PR 20-MAR-2001; 2001US-276947P.  
XX  
PA (ORTH ) ORTHO CLINICAL DIAGNOSTICS INC.  
XX  
PI Wan J, Wang Y;  
XX  
DR WPI: 2002-740862/80.  
XX  
PT New expression profile generated from primary, endothelial,  
PT epithelial, and muscle cell types, useful for identifying disease  
PT pathologies involving alterations of gene expression, e.g. cancer  
XX  
PS Disclosure; Page 641-642; 850pp; English.  
XX  
CC The invention relates to a gene expression profile comprising one or more  
CC genes (AB234889-AB235692) and generated from a cell type. The cell type

CC is a coronary artery endothelium, umbilical artery or vein endothelium,  
CC aortic endothelium, dermal microvascular endothelium, pulmonary artery  
CC endothelium, myometrium microvascular endothelium, keratinocyte  
CC epithelium, bronchial epithelium, mammary epithelium, prostate  
CC epithelium, renal cortical epithelium, renal proximal tubule epithelium,  
CC small airway epithelium, renal epithelium, umbilical artery smooth  
CC muscle, neonatal dermal fibroblast, pulmonary artery smooth muscle,  
CC dermal fibroblast, neural progenitor cells, skeletal muscle, astrocytes,  
CC aortic smooth muscle, mesangial cells, coronary artery smooth muscle,  
CC bronchial smooth muscle, uterine smooth muscle, lung fibroblast,  
CC osteoblasts or prostate stromal cell. The gene expression profile is used  
CC for determining the level of RNA expression for a sample, determining the  
CC phenotype of a cell and distinguishing cell types. The gene or a protein  
CC expression profile is useful in identifying disease pathologies  
CC involving alterations of gene expression. The assessment of expression  
CC profiles may provide meaningful information with respect to tumour type  
CC and stage, treatment methods, and prognosis. The gene or protein  
CC expression profile may also be used for creating microarrays. The  
CC microarray is useful for genetic and physical mapping of genomes, DNA  
CC sequencing, genetic or medical diagnosis, genotyping of organisms,  
CC confirming cell or tissue identifications and in identifying promising  
CC antibiotics, antiviral or antifungal agents.  
XX  
SQ Sequence 2167 BP; 537 A; 578 C; 454 G; 594 T; 4 other;

Query Match 16.1%; Score 164; DB 24; Length 2167;  
Best Local Similarity 71.2%; Pred. No. 2,2e-40;  
Matches 282; Conservative 0; Mismatches 90; Indels 24; Gaps 4;

QY 575 TCTTACAGTGTCTCTTGAAGCCGCCCTCCCTGCTGCTCCCTCCCAAGAGGGCGT 634  
DB 1780 TGTTCCTCAAGCCTTCTCCCTTCTTACACGCTGCAATGCTCCCAATTTGGGCA 1839  
QY 635 TGAGTCATATTTAACTGGCCATCCCTCAGAGTGTCTTAAGTCAAGTCTTCTTA 694  
DB 1840 TTAATTCATCTTTAACTGGTTCTGTAGTGGTAACTAGTAACTGTTCTTTAT 1899  
QY 695 GGACCCCTCTTAAAGCAATATGTCTGACCTGTACTTAAGATCTTCTGATATGC 754  
DB 1900 AGAACCCCTTCTACGAGCAATATGCTT-CTTGTATTAATAAATCTTCTGATATGC 1958  
QY 755 ATTGGAGATTTTGTGTAGATAGTAAAGTGCCTGCTTCTTACCTCTCTTATC 814  
DB 1959 ATTGAAGGTTTGTGATGATTAAGGCTTCTTCAATGATC-----TTATTCAG 2013  
QY 815 AGCTGACGTAGTCTCTCTGTTCTTCTAAGTGGGTAAATACAGTCTGCGGC 874  
DB 2014 AGCTAATATAGTCTTCTTCTTCTTCTAAGTGGGTAAATATCATATGTTGACG 2073  
QY 875 TTTACAGTTTAACTATTTAGATATTCGTA-----AACATCACTGT 918  
DB 2074 --TATAGTTTAAATATTTTATGATATCTTAACTAGAACCTTCTTAACATCACTGT 2131  
QY 919 CTTGCCAGATACCACTGTCATGTGATGTC 954  
DB 2132 CTTGCCAGATACCACTGTCATGTGATGTC 2167

RESULT 13  
ABL71132  
ID ABL71132 standard; cDNA; 285 BP.  
XX  
AC ABL71132;  
XX  
DT 14-MAY-2002 (first entry)  
XX  
DE Corn tassell-derived polynucleotide (cdps) SEQ ID NO:506.  
XX  
KW Corn; corn tassell-derived polynucleotide; cdps; hybrid breeding; CDPS;  
KW inheritance; characteristic; growth; development; disease resistance;  
KW environmental adaptability; quality; yield; molecular marker;  
KW multigene trait; plant breeding; corn tassell; gene; ss.  
XX

OS Zea mays.

XX US2001051335-A1.

XX 13-DEC-2001.

XX 16-APR-1999; 99US-0294093.

XX 21-APR-1998; 98US-082567P.

XX (LALG/) LALGUDI R V.

XX (ITOL/) ITO L Y.

XX (SHER/) SHERMAN B K.

XX Lalgudi RV, Ito LY, Sherman BK;

XX WPI; 2002-163647/21.

PT Novel purified corn tassel-derived polynucleotide useful for  
 PT determining altered gene expression, to recover regulatory elements and  
 PT breeding programs -

PS Claim 1; SEQ ID 506; 201pp; English.

XX The present sequence describes a purified corn tassel-derived  
 CC polynucleotide sequence (cdps) comprising a nucleic acid sequence  
 CC selected from those given in ABL70627 to ABL76833. The cdps sequences  
 CC encode corn tassel-derived polypeptides (CDPs). The cdps sequences (1)  
 CC can be used for determining altered gene expression, to recover  
 CC regulatory elements and to follow inheritance of desirable  
 CC characteristics through hybrid breeding programs. (1) are also useful  
 CC in the evaluation, and alteration of desired characteristics associated  
 CC with growth and development, disease resistance, environmental  
 CC adaptability, quality and yield, and as molecular markers for studying  
 CC inheritance of multiple traits in a plant breeding program. (1) can be  
 CC used to produce a tassel-specific profile of gene transcription, a  
 CC transcript image, to clone regulatory elements for use in transformation  
 CC vectors, to express a polypeptide, to identify, isolate or extend  
 CC identical or related corn tassel nucleic acid sequences from DNA  
 CC libraries, in nucleic acid hybridisation or amplification technologies,  
 CC as query sequences to determine homology of known sequences, as probe  
 CC for use in Southern or Northern hybridisation, and to identify the  
 CC presence of and/or to determine the degree of similarity between two  
 CC (or more) nucleic acid sequences.

XX Sequence 285 BP; 79 A; 46 C; 44 G; 115 T; 1 other;

Query Match 12.4%; Score 126.4; DB 24; Length 285;

Best Local Similarity 72.7%; Pred. No. 4.4e-29; Mismatches 66; Indels 6; Gaps 2;

Matches 192; Conservative 0; Mismatches 66; Indels 6; Gaps 2;

QY 658 TCCTACAGTGTCTTAACAGTGTCTTTCTTTAGACCCCTCTTAACGAGCAAT 717

DB 5 TTCTAGTAGTGTCTTAACAGTGTCTTTCTTTAGACCCCTCTTAACGAGCAAT 64

QY 718 ATGCTGACCTGTCTTAACAGTGTCTTTCTTTAGACCCCTCTTAACGAGCAAT 777

DB 65 ATGCTGACCTGTCTTTATTAACAGTGTCTTTCTTTAGACCCCTCTTAACGAGCAAT 123

QY 778 AGTAGAGTGTCTTTCTTTAGACCCCTCTTTAGACCCCTCTTTAGACCCCTCTTTAG 837

DB 124 AGTAGAGTGTCTTTCTTTAGACCCCTCTTTAGACCCCTCTTTAGACCCCTCTTTAG 178

QY 838 TTCTAGTAGTGTCTTAACAGTGTCTTTCTTTAGACCCCTCTTTAGACCCCTCTTTAG 897

DB 179 TTCTAGTAGTGTCTTAACAGTGTCTTTCTTTAGACCCCTCTTTAGACCCCTCTTTAG 238

QY 898 GATATCTGAAACATCACTGTCTT 921

DB 239 ATAATCTTAACATCACTGTCTT 262

RESULT 14

XX ABX43616/C

XX ID ABX43616 standard; CDNA; 416 BP.

XX AC ABX43616;

XX 21-FEB-2003 (first entry)

XX Bovine EST associated with lactation/muscle/fat deposition #8781.

XX Bovine: ss; EST: expressed sequence tag; lactation; LMPD:

XX muscle deposition; fat deposition; genome mapping; gene identification;

XX gene analysis; cattle breeding.

XX Bos Taurus.

XX US2002137139-A1.

XX 26-SEP-2002.

XX 24-SEP-2001; 2001US-0960352.

XX 12-JAN-1999; 99US-115707P.

XX 11-JAN-2000; 2000US-0480902.

XX (BYAT/) BYATT J C.

XX (MATH/) MATHALAGAN N.

XX (TAON/) TAO N.

XX (WARR/) WARREN W C.

XX Byatt JC, Mathalagan N, Tao N, Warren WC;

XX WPI; 2003-110599/10.

XX New nucleic acid associated with lactation, and muscle and fat  
 XX deposition, useful for genome mapping, gene identification and  
 XX analysis, cattle breeding, or for genetically improving cattle -

PS Claim 2; SEQ ID No 8781; 245pp; English.

XX The invention relates to a purified nucleic acid molecule associated with  
 CC lactation or muscle and fat deposition (designated LMPD), derived  
 CC from cattle, and the LMPD nucleic acid can specifically hybridise to a  
 CC second nucleic acid molecule comprising any of 15112 nucleotide  
 CC sequences, appearing as ABX436-ABX4947, or complements of them.  
 CC Also included are: (1) a transformed cell having a nucleic acid  
 CC comprising an LMPD nucleic acid linked to a promoter and a 3' non-  
 CC translated sequence that functions in the cell to cause termination of  
 CC transcription and addition of polyadenylated ribonucleotides to a 3' end  
 CC of the mRNA molecule; and (2) determining a level or pattern of a  
 CC molecule in a bovine cell or tissue comprising: (a) incubating a marker  
 CC nucleic acid (comprising any of the 15112 nucleic acid sequences or its  
 CC complement or fragment) with a complementary nucleic acid molecule  
 CC obtained from the bovine cell or tissue, where hybridisation between the  
 CC marker nucleic acid and the complementary nucleic acid permits the  
 CC detection of the molecule; and (b) detecting the level or pattern of the  
 CC complementary nucleic acid, where the detection of the complementary  
 CC nucleic acid is predictive of the level or pattern of the molecule.  
 CC The LMPD nucleic acid is used for determining a level or pattern  
 CC of a molecule in a bovine cell or tissue. It is useful for genome  
 CC mapping, gene identification and analysis, cattle breeding, preparation  
 CC of constructs for use in cattle gene expression, or for genetically  
 CC improving cattle. The present sequence is one of the 15112 bovine  
 CC LMPD EST (expressed sequence tag) nucleic acids.  
 CC Note: The present sequence was not shown in the specification but  
 CC was obtained in electronic format from the USPTO web site:  
 CC [seqdata.uspto.gov/sequence.html?docID=20020137139](http://seqdata.uspto.gov/sequence.html?docID=20020137139).

XX Sequence 416 BP; 140 A; 77 C; 101 G; 98 T; 0 other;

Query Match 12.1%; Score 123.2; DB 25; Length 416;

Best Local Similarity 78.2%; Pred. No. 5.4e-28; Mismatches 48; Indels 4; Gaps 3;



**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 25, 2003, 19:49:37 ; Search time 2115.56 seconds  
(without alignments)  
11683.734 Million cell updates/sec

Title: US-09-717-321A-15

Perfect score: 1017

Sequence: 1 cccctactctgtctcagatt.....ccttgggtctgtgaggttc 1017

Scoring table: IDENTITY\_NUC  
Gapop 10.0, Gapext 1.0

Searched: 22781392 seqs, 12152238056 residues

Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

EST:  
1: em\_estba:\*  
2: em\_esthum:\*  
3: em\_estin:\*  
4: em\_estmu:\*  
5: em\_estov:\*  
6: em\_estopl:\*  
7: em\_estro:\*  
8: em\_hic:\*  
9: gb\_est1:\*  
10: gb\_est2:\*  
11: gb\_hic:\*  
12: gb\_est3:\*  
13: gb\_est4:\*  
14: gb\_est5:\*  
15: em\_estfun:\*  
16: em\_estom:\*  
17: em\_gss\_hum:\*  
18: em\_gss\_inu:\*  
19: em\_gss\_pln:\*  
20: em\_gss\_vrt:\*  
21: em\_gss\_fun:\*  
22: em\_gss\_mam:\*  
23: em\_gss\_mus:\*  
24: em\_gss\_pro:\*  
25: em\_gss\_rtd:\*  
26: em\_gss\_phg:\*  
27: em\_gss\_vrt:\*  
28: gb\_gss1:\*  
29: gb\_gss2:\*

Pred. NO. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	678.2	66.7	1796	AK081613	Mus muscu
2	677.6	66.6	2269	AK034601	Mus muscu
3	677.6	66.6	2270	AK047969	Mus muscu
4	660.2	64.9	2308	AK076023	Mus muscu

5	594.2	58.4	978	13	B0514423	B0514423	AGENCOURT
6	562	55.3	826	14	CA318577	CA318577	UI-M-FW0-
7	558.2	54.9	751	13	B0177713	B0177713	UI-M-ERO-
8	552.2	54.3	960	13	B0511766	B0511766	AGENCOURT
9	551.2	54.2	557	10	B665146	B665146	DRAB2A12
10	537.4	52.8	793	14	CA319273	CA319273	UI-M-FW0-
11	535.2	52.6	897	12	B1646446	B1646446	603279026
12	528.6	52.0	714	12	B1851489	B1851489	603377691
13	528.6	52.0	915	13	B0931374	B0931374	AGENCOURT
14	526	51.7	545	9	AAB17948	AAB17948	UI-R-AO-a
15	526	51.7	639	13	B0201343	B0201343	UI-R-D01-
16	525.4	51.7	760	13	B0191985	B0191985	UI-R-DN1-
17	510.4	50.2	806	13	B0444803	B0444803	UI-M-ERO-
18	508.2	50.0	515	14	CD372599	CD372599	UI-M-G00-
19	499.4	49.1	644	14	CA339407	CA339407	NISC_Ly02
20	493.8	48.6	686	14	CB057853	CB057853	NISC_Ly09
21	490.8	48.3	672	12	B1692509	B1692509	603343041
22	485	47.7	969	10	BF783603	BF783603	602109609
23	483.4	47.5	680	14	CD353147	CD353147	UI-M-G10-
24	482.8	47.5	741	12	B1694684	B1694684	603347620
25	482.4	47.4	732	14	CB596051	CB596051	AGENCOURT
26	479	47.1	502	9	A1028951	A1028951	UI-R-C0-1
27	474.8	46.7	787	12	B0968782	B0968782	602837287
28	473.4	46.5	705	13	B0200122	B0200122	UI-R-EB1-
29	472.4	46.5	687	12	BM389059	BM389059	UI-R-D20-
30	462.4	45.5	796	12	B1735385	B1735385	603356871
31	458.4	45.1	588	14	CA553516	CA553516	COB73C09-
32	454.2	44.7	805	12	B1692090	B1692090	603311752
33	453	44.5	584	14	CB055806	CB055806	NISC_Ly09
34	450.8	44.3	667	12	B0042534	B0042534	UI-M-EMO-
35	449.2	44.2	938	13	B0928530	B0928530	AGENCOURT
36	442.8	43.5	447	10	B0663683	B0663683	DRAAAB04
37	438.8	43.1	671	12	BM946306	BM946306	UI-M-EMO-
38	438.2	43.1	704	12	BM951935	BM951935	UI-M-EMO-
39	427.2	42.0	538	12	BM035597	BM035597	1468B05.Y
40	426.6	41.9	550	10	BM022481	BM022481	1461A04.Y
41	426.2	41.9	950	10	BF168374	BF168374	601774359
42	426	41.9	556	12	B1655954	B1655954	603285490
43	425.4	41.8	575	10	BF659623	BF659623	u288905.x
44	423.8	41.7	640	13	B0780117	B0780117	UI-R-PP0-
45	422	41.5	702	13	B0176893	B0176893	UI-M-D12-

#### ALIGNMENTS

RESULT 1  
LOCUS AK081613 1796 bp mRNA linear HTC 05-DEC-2002  
DEFINITION Mus musculus 16 days embryo head cDNA, RIKEN full-length enriched library, clone:G130051H24, product:RMS-related C3 botulinum substrate 1, full insert sequence.

ACCESSION AK081613  
VERSION AK081613.1 GI:26349264  
KEYWORDS HTC; CAP trapper.  
SOURCE Mus musculus (house mouse)  
ORGANISM Mus musculus

REFERENCE  
1 Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Kono, H., Okazaki, Y., Muramatsu, M., and Hayashizaki, Y. High-efficiency full-length cDNA cloning by RT-PCR. *Nat. Biotechnol.* 17: 1273-1278 (1999)  
2 Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Kono, H., Okazaki, Y., Muramatsu, M., and Hayashizaki, Y. Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. *Genome Res.* 10 (10), 1617-1630 (2000)

TITLE JOURNAL MEDLINE PUBMED  
AUTHORS JOURNAL MEDLINE PUBMED  
10349636  
2  
Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Kono, H., Okazaki, Y., Muramatsu, M., and Hayashizaki, Y. Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. *Genome Res.* 10 (10), 1617-1630 (2000)  
20499374  
11042159

REFERENCE  
AUTHORS  
3  
Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P.,  
Konno, H., Akiyama, J., Nishi, K., Kitsuai, T., Tashiro, H., Itoh, M.,  
Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A.,  
Yamamoto, H., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K.,  
Fujisake, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Matsubara, J.,  
Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura, S., Kawai, J.,  
Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A., and Hayashizaki, Y.  
RIKEN Integrated sequence analysis (RISA) system -384-format  
sequencing pipeline with 384 multicapillary sequencer  
Genome Res. 10 (11), 1757-1771 (2000)

JOURNAL  
MEDLINE  
PUBMED  
20530913  
11076861

TITLE  
4  
Kawai, J., Shinagawa, A., Shibata, K., Yoshino, M., Itoh, M., Ishii, Y.,  
Aizawa, T., Hara, A., Fukunishi, Y., Konno, H., Adachi, J., Fukuda, S.,  
Aizawa, K., Izawa, M., Nishi, K., Kiyosawa, H., Kondo, S., Yamana, I.,  
Saito, T., Okazaki, Y., Gojobori, T., Bono, H., Kasukawa, T., Saito, R.,  
Kadota, K., Matsuda, H., Ashburner, M., Batalov, S., Casavant, T.,  
Fleischmann, W., Gaasterland, T., Gissi, C., King, B., Kochiwa, H.,  
Kuehl, P., Lewis, S., Matsuo, Y., Nikaido, I., Pesole, G.,  
Quackenbush, J., Schriml, L.M., Staudli, F., Suzuki, R., Tomita, M.,  
Wagner, L., Washio, T., Sakai, K., Okido, T., Furuno, M., Aono, H.,  
Baldarelli, R., Barsh, G., Blake, J., Boffelli, D., Bojunga, N.,  
Carninci, P., de Bonaldo, M.F., Brownstein, M.J., Bult, C.,  
Fletcher, C., Fujita, M., Cariboldi, M., Gustincich, S., Hill, D.,  
Hofmann, M., Hume, D.A., Kamiya, M., Lee, N.H., Lyons, P.,  
Marchionni, L., Mashima, J., Mazzarelli, J., Mombaerts, P., Nordone, P.,  
Riing, B., Ringwald, M., Rodriguez, I., Sakamoto, N., Sasaki, H.,  
Sato, K., Schonbach, C., Seya, T., Shibata, Y., Storch, K.F., Suzuki, H.,  
Toyo-Oka, K., Wang, K.H., Weitz, C., Whitaker, C., Willing, L.,  
Wynshaw-Boris, A., Yoshida, K., Hasegawa, Y., Kawai, H., Kohcuki, S.,  
and Hayashizaki, Y.  
Functional annotation of a full-length mouse cDNA collection  
Nature 409 (6821), 685-690 (2001)

JOURNAL  
MEDLINE  
PUBMED  
21085660  
11217851

TITLE  
5  
The FANTOM Consortium and the RIKEN Genome Exploration Research  
Group Phase I & II Team.  
Analysis of the mouse transcriptome based on functional annotation  
of 60,770 full-length cDNAs  
Nature 420, 563-573 (2002)

JOURNAL  
REFERENCE  
AUTHORS  
6 (bases 1 to 1796)  
Adachi, J., Aizawa, K., Akimura, T., Arikawa, T., Bono, H., Carninci, P.,  
Fukuda, S., Furuno, M., Hanagaki, T., Hara, A., Hashizume, W.,  
Hayashida, K., Hayatsu, N., Hiramoto, K., Hiraoaka, T., Hirozane, T.,  
Hori, F., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kasukawa, T.,  
Kato, H., Kawai, J., Kojima, Y., Kondo, S., Konno, H., Kouda, M.,  
Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M.,  
Nakamura, Y., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Ohsato, N.,  
Okazaki, Y., Saito, R., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N.,  
Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, T.,  
Sogabe, T., Tagami, M., Tagawa, A., Takahashi, F., Takaku-Kahira, S.,  
Takeda, T., Tanaka, T., Tomaru, A., Toya, T., Yasunishi, A.,  
Muramatsu, M., and Hayashizaki, Y.  
Direct Submission

TITLE  
JOURNAL  
COMMENT  
Submitted (16-APR-2002) Yoshihide Hayashizaki, The Institute of  
Physical and Chemical Research (RIKEN), Laboratory for Genome  
Exploration Research Group, RIKEN Genomic Sciences Center (GSC),  
RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,  
Kanagawa 230-0045, Japan (E-mail: genome-res@gs.c.riken.go.jp,  
URL: http://genome.gsc.riken.go.jp/, Tel: 81-45-503-9222,  
Fax: 81-45-503-9216)  
cDNA library was prepared and sequenced in Mouse Genome  
Encyclopedia Project of Genome Exploration Research Group in Riken  
Genomic Sciences Center and Genome Science Laboratory in RIKEN  
Division of Experimental Animal Research in Riken contributed to  
prepare mouse tissues.  
Please visit our web site for further details.  
URL: http://genome.gsc.riken.go.jp/  
URL: http://fantom.gsc.riken.go.jp/  
Location/Qualifiers

FEATURES

source  
1. 1796  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/strain="C57BL/6J"  
/db\_xref="FANTOM:DB:C130051H24"  
/db\_xref="taxon:10090"  
/clone="C130051H24"  
/issue\_type="head"  
/clone\_lib="RIKEN full-length enriched mouse cDNA library"  
/dev\_stage="16 days embryo"  
<1. 304  
/note="unnamed protein product; PAS-related C3 botulinum  
substrate 1 (MGI:97845, GB|K011072, evidence: BLASTN,  
100%, match=987)  
putative"  
/codon\_start=2  
/protein\_id="BAC38272.1"  
/db\_xref="GI:26349265"  
/translation="VRAKYPPEYRHHCPPIIIVTGTDLDRDDKTIKLEKILTP  
ITVQGLAAKEIGAVYDECSALVGRGLKTVFDEAIRAVLCPPEPKRRKRCLL"  
1781. 1786  
/note="putative"  
1796  
polya\_signal  
polya\_site  
BASE COUNT 470 a 413 c 371 g 542 t  
ORIGIN  
Query Match 66.7%; Score 678.2; DB 11; Length 1796;  
Best Local Similarity 85.3%; Pred. No. 1.7e-183;  
Matches 871; Conservative 0; Mismatches 118; Indels 32; gaps 9;  
QY 1 CCCCTATCTCTGCTCAGATTAAAGATTGCCAAATTCCTGTGTAAGTTCGCTTGG 60  
DB 548 CCCCATTCTCTGCTCAGATTAAAGATTGCCAAATTCCTGTGTAAGTTCGCTTGG 607  
QY 61 CTGAGAACAGTAAAGACCTAAGCTGTGAGAGA---CTTGTGCTTAAAGAGCTGACG 117  
DB 608 CTGAGAACAGTAAAGACCTAAGCTGTGAGAGACTTCTGTGTAAGAGCCGACG 667  
QY 118 TTCTGGGCTCAGGGGTCAGACACCTCCCTGAGTCCACAGCCGTGACACAGACG 177  
DB 668 TTCTGGGCTCAGGGGTCAGACACCTCCCTGAGTCCACAGCCGTGACACAGACG 727  
QY 178 TCCTTAATGACAGCGTCCATGTAAGCAGCTTAATCTATGACCCATGCTATTACG 237  
DB 728 TCCTTAATGACAGCGTCCATGTAAGCAGCTTAATCTATGACCCATGCTATTACG 787  
QY 238 AACCTTGTAAGTACGTCAC---GATGGGTGAACAGCTGCTGCTTGTGATTCATAG 294  
DB 788 AACCTTGTAAGTACGTCACAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGT 847  
QY 295 AGTTCTCTAAATACAGCTGACCGGCTTCTGAGGCTTGAACAGAACTGCTCCTG 354  
DB 848 AGTTCTCTAAATACAGCTGACCGGCTTCTGAGGCTTGAACAGAACTGCTCCTG 907  
QY 355 TGTTCCTCTAAGAAATATCTGTTCTGTTCTGTTCTGTTCTGTTCTGTTCTG 414  
DB 908 TGTTCCTCTAAGAAATATCTGTTCTGTTCTGTTCTGTTCTGTTCTGTTCTG 965  
QY 415 ACACGAGCTCATGAAGAGACAGACAGATTTTGAATATGAAGTGAAGTGAAGT 474  
DB 966 ACACGAGCTCATGAAGAGACAGACAGATTTTGAATATGAAGTGAAGTGAAGT 1025  
QY 475 ACACGAGCTCATGAAGAGATTAATCAATGAATTAAGTGAAGTGAAGTGAAGT 534  
DB 1026 ACACGAGCTCATGAAGAGATTAATCAATGAATTAAGTGAAGTGAAGTGAAGT 1085  
QY 535 ACAGTTAATTTCTGTAACAGTGAATGAATGAATGAATGAATGAATGAATGA 589  
DB 1086 ACAGTTAATTTCTGTAACAGTGAATGAATGAATGAATGAATGAATGAATGA 1145  
QY 590 -----TCTGAGCCCGCTTCCCGCTGCTCCCTCCGATGAGGCGTGAATCCA 642  
|| | |||| | || |||| |||| |||| |||| |||| |||| |||| |||| ||||



Db	1146	CATGCCCTTCGGCCGCCCCACACCCCTGCACATGCGCGTCCCTCCACAGATGGGGGGCTTGAGTCCCA	1205
Qy	643	TATTTAAACCTGGCCATCTCTCAGAGTGTGTAACCTAGCAGATGCTTTTCTTAGAGACCCC	702
Db	1206	TATTTAAACCTGGCCACCCCTGCACAGT-T-CTAATTAAGCAAGGCTTTCTTTA-GAACCC	1263
Qy	703	TTCTTAAAGCAATATGTCTGACCTGTGACTGACTATTAAGATCTTCTTGATTAATGCA-----T	756
Db	1264	TTCTTAAAGCAATATGTCTGACCTGTGACTGACTATTAAGATCTTCTTGATTAAGCAAGATTT	1323
Qy	757	TCGAGATTTTTTTGGTAGATAGTAGAAGTGCCTCTGTTTTCACCTTCCTTACTACAG	816
Db	1324	TTTTTTTTTTTTTTGGTAGCTCAGTAGAAGTGCCTCTGTTTTCACCTTACTTACTACAG	1383
Qy	817	-CTACAGAGTCTCTCCCTGCTTTTCTAGTAACTGGGGTGGTGAATATCAGTGGCGGGCT	875
Db	1384	AGCTAGTATGCTCTCTAGTATTCCTAGCAACTAGGCTGTCAAAATCATGTGTGGACAGT	1443
Qy	876	TTACAGTTTTTAAACTATTTAGATATTTCTGAACAATCACAGTGTGGCCAGAGTACCAAC	935
Db	1444	TTAGGTTTTTAAACATNTTTAGATATTTCTTAAACATATGAACCTTCTTAAACAGTACCAAC	1503
Qy	936	ACTGTCAATGATTTGATGCGCGCCCTCTAGACCTCACCAAGCGGAGACATGCTTCCGG	995
Db	1504	ACTGTCAATGATTTGATGCTGCTGCTCTAGACCTCACCAAGCGGAGACATGCTTCCGT	1560
Qy	996	T 996	
Db	1561	T 1561	

ARKAWA, T., HARA, A., FUKUNISHI, Y., KONNO, H., AAGCHI, J., FUKUDA, S., AITAKE, K., IRAWA, M., NISHI, K., KIYOSWA, H., KONDO, S., YAMANAKA, I., SATTO, T., OKAZAKI, Y., GOJOORI, T., BONO, H., KASUKAWA, T., SATO, R., KADOTA, K., MATSUDA, H., ASHBURNER, M., BATALOV, S., CASAVANT, T., FLEISCHMAN, W., GAESTERLAND, T., GISSI, C., KING, B., KOCHIWA, H., KUEHL, P., LEVITS, S., MATSUO, Y., NAKAI, I., PEEBLE, G., QUACKENBUSH, J., SCHIRMLI, L. M., STABILI, F., SUZUKI, R., TOMITA, M., WAGNER, L., WASHIO, T., SAKAI, K., OKIDO, T., FURUNO, M., AONO, H., BALDARELLI, R., BERSH, G., BLAKE, J., BROWNSTEIN, M. J., BULT, C., CARLINI, P., DE DONALDO, M. F., GUSTINICH, S., HILL, D., FLETCHER, C., FUJITA, M., GARIBOLDI, M., GUSTINICH, S., HILL, D., HOTMANN, M., HUME, D. A., KAMIYA, M., LEE, N. H., LYONS, P., MACHIONNI, L., MASHIMA, J., MAZZARELLI, J., MOMBERTS, P., NORDONE, P., RING, B., RINGWALD, M., RODRIGUEZ, I., SAKAMOTO, N., SASAKI, H., SATO, K., SCHONBACH, C., SEYA, T., SHIBATA, Y., STORCH, K. F., SUZUKI, H., TOYO-OKA, K., WANG, K. H., WELTZ, C., WHITTAKER, C., WILMING, L., WYSHAW-BORIS, A. A., YOSHIDA, K., HASEGAWA, Y., KAWAJI, H., KOHTSUKI, S. and HAYASHIZAKI, Y.	Functional annotation of a full-length mouse cDNA collection Nature 409 (6821), 685-690 (2001)
TITLE JOURNAL MEDLINE PUBMED 11217851	
5	
REFERENCES	
TITLE	
JOURNAL	
REFERENCE	

RESULT 2	
AKO34601	
LOCUS	AKO34601
DEFINITION	Mus musculus 12 days embryo embryonic body between diaphragm region and neck cDNA, RIKEN full-length enriched library, clone:9430013G17
ACCESSION	AKO34601
VERSION	AKO34601.1 GI:26330056
KEYWORDS	HTC; CAP trapper.
SOURCE	Mus musculus (house mouse)
ORGANISM	Mus musculus
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
AUTHORS	1 Carninci, P. and Hayashizaki, Y.
TITLE	High-efficiency full-length cDNA cloning
JOURNAL	Meth. Enzymol. 303, 19-44 (1999)
PUBMED	99279253
REFERENCE	2 10349636
AUTHORS	3 Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
TITLE	Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes
JOURNAL	Genome Res. 10 (10), 1617-1630 (2000)
PUBMED	20499374
REFERENCE	4 11042159
AUTHORS	5 Shobha, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P., Konno, H., Akiyama, J., Ishii, K., Kitasuna, T., Tashiro, H., Itoh, M., Suni, N., Ishii, Y., Nakamura, S., Hazama, M., Nishize, T., Harada, A., Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K., Fujiwake, S., Inoue, K., Togawa, Y., Iwawa, M., Ohara, E., Watabiki, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsunura, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.
TITLE	RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer
JOURNAL	Genome Res. 10 (11), 1757-1771 (2000)
PUBMED	20530913
REFERENCE	6 11076661
AUTHORS	7 Kawai, J., Shinagawa, A., Shibata, K., Yoshino, M., Itoh, M., Ishii, Y.,

	Hayashida,K., Hayatsu,N., Hiramoto,K., Hiroka,T., Hitozono,T., Horii,P., Imocani,K., Ishii,Y., Itoh,M., Kagawa,I., Kakuwa,T., Katoh,H., Kawaji,J., Kojima,Y., Kondo,S., Konno,H., Kouda,M., Koya,S., Kurihara,C., Matsuyama,T., Miyazaki,A., Murata,M., Nakamura,M., Nishi,R., Nomura,K., Nunazaki,R., Ohno,M., Ohsato,N., Okazaki,Y., Saito,R., Saitoh,H., Sakai,C., Sakai,K., Sakazume,N., Sano,H., Sasakl,D., Shibata,K., Shingawa,A., Shiraki,T., Sugabe,Y., Tagami,M., Tagawa,A., Takahashi,F., Takeku-Akahito,S., Takeda,Y., Tanaka,T., Tomaru,A., Toya,T., Yasunishi,A., Muramatsu,M. and Hayashizaki,Y.
TITLE	Direct Submission
JOURNAL	Submitted (16-Jul-2001) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration and Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama Kanagawa 230-0045, Japan (E-mail:genome-resgsc@riken.go.jp, FAX:81-45-503-9216)
COMMENT	cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. Please visit our web site for further details. URL:http://genome.gsc.riken.go.jp/ URL:http://fantom.gsc.riken.go.jp/ Location/Qualifiers 1..2269
FEATURES	/organism="Mus musculus" /mol_type="mRNA" /strain="C57BL/6J" /db_xref="FANTOM_DB:9430013617" /db_xref="taxon:10090" /clone="9430013617" /tissue-type="embryonic body between diaphragm region and neck"
SOURCE	/clone.lib="RIKEN full-length enriched mouse cDNA library" /dev_stage="12 days embryo" 198..776 /note="(unnamed protein product; Ras-related C3 botulinum substrate 1 [MGDI/MGI:97845, GBI/AKO11072, evidence: BLASTN 100%, match=987])
CDS	



Wynshaw-Boris, A., Yoshida, K., Hasegawa, Y., Kawaji, H., Kohlsunki, S., and Hayashizaki, Y.  
Functional annotation of a full-length mouse cDNA collection  
JOURNAL  
MEDLINE  
21085660  
PUBMED  
11217851  
REFERENCE  
AUTHORS  
5  
TITLE  
The FANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team.  
Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs  
JOURNAL  
REFERENCE  
AUTHORS  
6 (bases 1 to 2270)  
Fukuda, S., Alizawa, K., Akimura, T., Arakawa, T., Bono, H., Carninci, P., Adachi, J., Furuno, M., Hanagaki, T., Hara, A., Hashizume, M., Hayashida, K., Hayatsu, N., Hiramoto, K., Hirooka, T., Hirozane, T., Horii, P., Imotanli, K., Ishii, Y., Itoh, M., Kagawa, I., Kasukawa, T., Katoh, H., Kawai, J., Kojima, Y., Kondo, S., Konno, H., Kouda, M., Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M., Nakamura, M., Nishii, K., Nomura, K., Numazaki, R., Ohno, M., Ohsato, N., Okazaki, Y., Saito, R., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N., Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, T., Sogabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takaku-Akahira, S., Takeda, Y., Tanaka, T., Tomaru, A., Toya, T., Yasunishi, A., Muramatsu, M., and Hayashizaki, Y.  
Direct Submission  
Submitted (16-JUL-2001) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail: genome-res@gs.c.riken.go.jp, URL: http://genome.gsc.riken.go.jp/, Tel: 81-45-503-9222, Fax: 81-45-503-9216)  
CDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.  
Please visit our web site for further details.  
URL: http://genome.gsc.riken.go.jp/  
URL: http://fantom.gsc.riken.go.jp/.  
FEATURES  
source  
Location/Qualifiers  
1..2270  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/strain="C57BL/6J"  
/db\_xref="FANTOM:DB:C130025F01"  
/db\_xref="taxon:10090"  
/clone="C130025F01"  
/tissue\_type="head"  
/clone\_lib="RIKEN full-length enriched mouse cDNA library"  
/dev\_stage="16 days embryo"  
200..778  
/note="unnamed protein product, Ras-related G3 botulinum substrate 1 (MGI:97845, GBAK01072, evidence: BLASTN, 100%, match=987)  
putative"  
/codon\_start=1  
/protein\_id="BAC3203.1"  
/db\_xref="GI:26339064"  
/translation="MQAIRCYVGDGANGKTCILISTYTNAPFGYITPTVDNSANVWVGGPVLGLMDRAGDEYDLRPLSPQTDVPLICSLVSPSEFVNRKWPVEVHHCNPTPIILVGTDLKDLDNDKDTLEKLEKLTPTIPYQGLAMAEIGAVYLLCSALTORGLKTFVDEAIRAVLCPPPVKRRKRKCLL"  
2254..2259  
polyA\_signal  
/note="putative"  
2270  
BASE COUNT 554 a 569 c 503 g 644 t  
ORIGIN  
Query Match 66.6%; Score 677.6; DB 11; Length 2270;  
Best Local Similarity 85.3%; Pred. No. 2.9e-183;

Matches 870; Conservative 0; Mismatches 119; Indels 31; Gaps 9;  
QY 1 CCCCTATCTCTGTCAGATTAGAAATGCGAAATACCTGTGACACTAGTGGCTGTG 60  
DB 1022 CCCCATCTCTGCTAGATTAAATGCGAAATACCTGTGACACTAGTGGCTGTG 1081  
QY 61 CTGAGACACGTAACTAGCTGTGAGAGA---CTTTGTCTTAAGAACATGCGAGC 117  
DB 1082 CTGAGACACCTTAACACACTAAACTCTGTGAGAGACTCTGTGTGAAGAACCGCAGC 1141  
QY 118 TTCTGGGCTCAGGGTGCAGACCCCTCCGTAAGTCCCAAGCCGTGTGACAGCAGCACC 177  
DB 1142 TTCTGAGCCAGGGGTGAGACCCCTCAGAGTTCCAGACCTGTCAGACAGCAGCAGCC 1201  
QY 178 TCCTTAATGACACGGTGCATGATAGGACACCTGTACTATACGCCATGCGCATTAAGT 237  
DB 1202 TCCTTAATGACACGGTGCATGATAGGACACCTGTACTATACGCCATGCGCATTAAGT 1261  
QY 238 AACCTTGTACTGTAGTGCAC---GATGGGTGTAAACAGCTGCTTTGATTCATTAAGT 294  
DB 1262 AACCTTGTACTGTAGTGCACGTAGTATGATGATGATGATGATGATGATGATGATG 1321  
QY 295 AGTTCCTTAATTAACAGCTGACCGGCTTGTGACAGCTTTGAAACGAACCTGCTGCTG 354  
DB 1322 AGTTCCTTAATTAACAGCTGACCGGCTTGTGACAGCTTTGAAACGAACCTGCTGCTG 1381  
QY 355 TGTTCCTCTTAACGAATATCTGTTCTAGTGTGAGGTGCTGGGTGAGTGTGTA 414  
DB 1382 TGTTCCTCTTAACGAATATCTGTTCTAGTGTGAGGTGCTGGGTGAGTGTGTA 1439  
QY 415 ACACAGCGTCATCAAAAGAGACAGACAGTATTTTGTACTAATATGAATAGATTAATTT 474  
DB 1440 ACACAGCGTCATCAAAAGAGAGACAGACAGTATTTTGTACTAATATGAATAGATTAATTT 1499  
QY 475 ACACGACATGTGACATGAGATTAATCACTAATTAAGTGTGACGGTAAAGCTTTTAA 534  
DB 1500 ACACGACATGTGACATGAGATTAATCACTAATTAAGTGTGACGGTAAAGCTTTTAA 1559  
QY 535 ACGGTAATTTCTGCAACAGTAGATGACAAATGGCCGATCTTTCAGTGTCTC----- 589  
DB 1560 ACGGTAATTTCTGCAACAGTAGATGACAAATGGCCGATCTTTCAGTGTCTC----- 1619  
QY 590 -----TCTTGAGCCGCCCTTCCCTGCTGCTGCTCCGCCAGATGGGGCGTTGATCCA 642  
DB 1620 CATCCCTTCGGGCCGCCACCCTCCAGTGCCTGCCCTCCAGATGGGGCGTTGATCCA 1679  
QY 643 TATTTAACTGGCCATCTCTACAGTGTGTAAGTGTGTAAGTGTGTAAGTGTGTAAGT 702  
DB 1680 TATTTAACTGGCCATCTCTACAGTGTGTAAGTGTGTAAGTGTGTAAGTGTGTAAGT 1737  
QY 703 TTCTTAACGAGCAATATGCTGACCTGTACTATAGATCTTTCGATTAATGCA-----TT 757  
DB 1738 TTCTTAACGAGCAATATGCTGACCTGTACTATAGATCTTTCGATTAATGCA-----TT 1797  
QY 758 CGGAGATTTTTTGGTAGATAGTAGAGTGGTGTCTGTTTACCTCTCTTACTACG- 816  
DB 1798 TTTGTTTGTGTGACTAGTAGAGTGGTGTCTGTTTACCTCTCTTACTACG- 1857  
QY 817 CTGACTAGTGTGCTCCCTGCTGTTTCTTACTAAGTGGGTGTAAATACAGTGTGCTG 876  
DB 1858 GCTACTAGTGTGCTCCCTGCTGTTTCTTACTAAGTGGGTGTAAATACAGTGTGCTG 1917  
QY 877 TACAGTTTAACTAATTTTGAATATTTGAAACATCACTGCTGTGCGAGATGACACA 936  
DB 1918 TACGTTTAACTAATTTTGAATATTTGAAACATCACTGCTGTGCGAGATGACACA 1977  
QY 937 CTGTCATGTGATGATGCGGCCCTTGAACCTCAGCCAGCGGACAGCATGCTTCGGT 996  
DB 1978 CTGTCATGTGATGATGCGGCCCTTGAACCTCAGCCAGCGGACAGCATGCTTCGGT 2034  
RESULT 4  
AK076023

LOCUS AK076023 2308 bp mRNA linear HTC 07-DEC-2002  
DEFINITION Mus musculus 10 days embryo whole body cDNA, RIKEN full-length  
subtracted library, clone:261010016 product:RAS-related C3 botulinum  
substrate 1, full insert sequence.  
ACCESSION AK076023  
VERSION AK076023.1 GI:26344957  
KEYWORDS HTC; CAP trapper.  
SOURCE Mus musculus (house mouse)  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
REFERENCE 1  
AUTHORS Carninci, P. and Hayashizaki, Y.  
TITLE High-efficiency full-length cDNA cloning  
JOURNAL Meth. Enzymol. 303, 19-44 (1999)  
MEDLINE 99279253  
PubMed 10349636  
REFERENCE 2  
AUTHORS Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,  
Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.  
TITLE Normalization and subtraction of cap-trapper-selected cDNAs to  
prepare full-length cDNA libraries for rapid discovery of new genes  
JOURNAL Genome Res. 10 (10), 1617-1630 (2000)  
MEDLINE 20499374  
PubMed 11042159  
REFERENCE 3  
AUTHORS Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P.,  
Konno, H., Akiyama, J., Nishi, K., Kitsuami, T., Tashiro, H., Itoh, M.,  
Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Tashiro, H., Harada, A.,  
Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K.,  
Fujisake, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Matsui, M.,  
Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura, S., Kawai, T.,  
Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.  
TITLE RIKEN integrated sequence analysis (RISA) system--384-format  
sequencing pipeline with 384 multicapillary sequencer  
JOURNAL Genome Res. 10 (11), 1757-1771 (2000)  
MEDLINE 20530913  
PubMed 11076861  
REFERENCE 4  
AUTHORS Kawai, J., Shingawa, A., Shibata, K., Yoshino, M., Itoh, M., Ishii, Y.,  
Arai, T., Hara, A., Fukunishi, Y., Konno, H., Adachi, J., Fukuda, S.,  
Aizawa, K., Izawa, M., Nishi, K., Kiyosawa, H., Kondo, S., Yamana, I.,  
Saito, T., Okazaki, Y., Gojobori, T., Bono, H., Kasukawa, T., Saito, R.,  
Kadoya, K., Matsuda, H., Ashburner, M., Batalov, S., Casavant, T.,  
Fleischmann, W., Gaasterland, T., Gissi, C., King, B., Kochiwa, H.,  
Kuehl, P., Lewis, S., Matsuo, Y., Nikaido, I., Pesole, G.,  
Wagner, L., Washio, T., Sakai, K., Okido, T., Suzuki, R., Tomita, M.,  
Baldarelli, R., Barsh, G., Blake, J., Boffelli, D., Bolunga, N.,  
Carninci, P., de Bonaldo, M. F., Brownstein, M. J., Bult, C.,  
Fletcher, C., Fujita, M., Gariboldi, M., Gustincich, S., Hill, D.,  
Hofmann, M., Hume, D. A., Kamiya, M., Lee, N. H., Lyons, P.,  
Marchionni, L., Mashima, J., Mazzarelli, J., Mombaerts, P., Nordone, P.,  
Rang, B., Ringwald, M., Rodriguez, I., Sakamoto, N., Sasaki, H.,  
Sato, K., Schonbach, C., Seya, T., Shibata, Y., Storch, K. F., Suzuki, H.,  
Toyo-oka, K., Wang, K. H., Weitz, C., Whitaker, C., Wilmink, L.,  
Wynshaw-Boris, A., Yoshida, K., Hasegawa, Y., Kawai, H., Kohsaki, S.,  
and Hayashizaki, Y.  
TITLE Functional annotation of a full-length mouse cDNA collection  
JOURNAL Nature 409 (6821), 685-690 (2001)  
MEDLINE 21085660  
PubMed 11217851  
REFERENCE 5  
AUTHORS The FANTOM Consortium and the RIKEN Genome Exploration Research  
Group Phase I & II Team.  
TITLE Analysis of the mouse transcriptome based on functional annotation  
of 60,770 full-length cDNAs  
JOURNAL Nature 420, 563-573 (2002)  
MEDLINE 12121851  
PubMed 12121851  
REFERENCE 6 (phases 1 to 2308)  
AUTHORS Adachi, J., Aizawa, K., Akahira, S., Akimura, T., Aono, H., Arai, A.,  
Arai, T., Bono, H., Carninci, P., Fukuda, S., Fukunishi, Y.,  
Furuno, M., Hanagaki, T., Hara, A., Hayatsu, N., Himeno, K.,  
Hiroaka, T., Hori, F., Imotani, K., Ishii, Y., Itoh, M., Izawa, M.,

kasukawa, T., Kato, H., Kawai, J., Kojima, Y., Konno, H., Kouda, M.,  
Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Nishi, K.,  
Nomura, K., Numazaki, R., Ohno, M., Okazaki, Y., Okido, T., Owa, C.,  
Saito, H., Saito, R., Sakai, C., Sakai, K., Sano, H., Sasaki, D.,  
Shibata, K., Shibata, Y., Shingawa, A., Shiraki, T., Sogabe, Y.,  
Suzuki, H., Tagami, M., Tagawa, A., Takahashi, F., Tanaka, T.,  
Tejima, Y., Toya, T., Yamamura, T., Yamana, I., Yasunishi, A.,  
Yoshida, K., Yoshino, M., Muramatsu, M. and Hayashizaki, Y.  
TITLE Direct Submission  
JOURNAL Submitted (16-APR-2002) Yoshihide Hayashizaki, The Institute of  
Physical and Chemical Research (RIKEN), Laboratory for Genome  
Exploration Research Group, RIKEN Genomic Sciences Center (GSC),  
RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,  
Kanagawa 230-0045, Japan (E-mail: genome-res@gs.riken.go.jp,  
URL: http://genome.gsc.riken.go.jp/, Tel: 81-45-503-9222,  
Fax: 81-45-503-9216)  
COMMENT cDNA library was prepared and sequenced in Mouse Genome  
Encyclopedia Project of Genome Exploration Research Group in Riken  
Genomic Sciences Center and Genome Science Laboratory in Riken.  
Division of Experimental Animal Research in Riken contributed to  
prepare mouse tissues.  
Please visit our web site for further details.  
URL: http://genome.gsc.riken.go.jp/  
URL: http://fantom.gsc.riken.go.jp/.  
FEATURES  
source  
1..2308  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/strain="C57BL/6J"  
/db\_xref="FANTOM,DB:261010016"  
/db\_xref="MG:1898616"  
/db\_xref="taxon:10090"  
/clone="261010016"  
/tissue\_type="whole body"  
/clone\_1kb="RIKEN full-length enriched mouse cDNA library"  
/dev\_stage="10 days embryo"  
239..817  
/note="unamed protein product; RAS-related C3 botulinum  
substrate 1 (MSD) MG1:97845, GB|AK011072, evidence: BLASTN,  
100%, match=987)  
putative"  
/codon\_start=1  
/protein\_id="BAC36128.1"  
/db\_xref="GI:26344958"  
/db\_xref="MG:197845"  
/translation="MQAIRCVVGDGAVKTCLLISYTNAPGEYIPYFDNYSNV  
MVDGKPVNLGIMDTAGQEDYRLRPLSPQTVFLICFSIVSAFENRARKYPEVR  
HRCPPNPIIVGKTLRDRDKDRIETKIKLPIYVPGGLAAKRIKAVKYLECSAL  
TORGLKTVDEAIRVALCPPEPKRRKRCULL"  
2293..2298  
/note="putative"  
polyA\_signal  
2293..2298  
polyA\_site  
2308  
/note="putative"  
BASE COUNT 559 a 580 c 518 g 650 t 1 others  
ORIGIN  
Query Match 64.9%; Score 660.2; DB 11; Length 2308;  
Best Local Similarity 84.8%; Pred. No. 3e-176;  
Matches 866; Conservative 0; Mismatches 123; Indels 32; Gaps 10;  
QY 1 CCCCTATTCCTGTCGATTAAGATTGCCAAATACCTGCACTAAGTTCGCTGTC 60  
DB 1061 CCCCATCTTGTGTCGATTAAGATTGCCAAATACCTGCACTAAGTTCGCTGTC 1120  
QY 61 CTGAGAACAGTAAGCACTAAGCTGTGAGAGA-CTTTGTCTCTTAAGAAAGCTCAGC 117  
DB 1121 CTGAGAACAGTAAGCACTAAGCTGTGAGAGACTTGTGCTGAAGAAAGCCAGC 1180  
QY 118 TTCTGG- GCTCAGGGGTGACAGCCCTCCGTAAGACCCGCTGACAGCAGCAGC 176  
DB 1181 TTCTGGAGGAGGAGGGGTGACAGCCCTCCGTAAGACCCGCTGACAGCAGCAGC 1240  
QY 177 CTCCTTAATGACACGCTGCCATGTAAGCAGCACTGAATTATAGCCCATGCTATTAGC 236

```

Db      1241 CTCCTTAATGACAGCTGCCATGTAACACACCTGTAATTAACGCCATGTCATTACG 1300
OY      237 TAACCTGTACTAGTACGTCAC---GATGGGTGTAACACTGTGCTTGTATTCATAGT 293
Db      1301 TAACCTGTACTAGTACGTCACAGTATAGTGTACACTGACCTGACCTGTTGATTCATAGT 1360
OY      294 GAGTCTCTAAATAACAGCTGACCGGCTTCTGACAGCTTGAACAGAACTGTGGCTCT 353
Db      1361 GAGTTTCTAAAGACCGACGACAGCTTGTGCACTTTGACAACTGTGGTCTCT 1420
OY      354 GTGTGCTCTGACAGAGTATTTCTGTTCTTAAGTCTGGGTGCTGCTGGAGTGTGA 413
Db      1421 GTGTGCTCTGACAGAGTATTTCTGTTCTTAAGTCTGGGTGCTGCTGGAGTGTGA 1478
OY      414 AACAGAGTATCAAAAGAGACAGACAGTATTTGACTAATATGAAATAGAGTATTT 473
Db      1479 AACAGAGTATCAAAAGAGACAGACAGTATTTGACTAATATGAAATAGAGTATTT 1538
OY      474 TACACTACATTTGACATGAGTAAATTCATGATGAAATGATGACGCGTAAAGCTTTT 533
Db      1539 TACACTACATTTGACATGAGTAAATTCATGATGAAATGATGACGCGTAAAGCTTTT 1598
OY      534 AACGTTAATTTCTGTCAAACAGTATGATGACAAATGCGCATTTATGATGCTC--- 589
Db      1599 AACGTTAATTTCTGTCAAACAGTATGATGACAAATGCGCATTTATGATGCTC--- 1658
OY      590 -----TCTGAGCCCCCTTCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 641
Db      1659 GCATCCCTCGCGGCCCCACCCCTCCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1718
OY      642 ATATTTAACTGCGCATCTCCACAGTGTCTAATGAGTAAAGTCTTTTCTTAAAGACCC 701
Db      1719 ATATTTAACTGCGCATCTCCACAGTGTCTAATGAGTAAAGTCTTTTCTTAAAGACCC 1776
OY      702 CTTCCTTAACGAGCAATATGCTGACCTGTACTATAGATCTTTGATATGCA-----T 756
Db      1777 CTTCCTTAACGAGCAATATGCTGACCTGTACTATAGATCTTTGATATGCA-----T 1836
OY      757 TCGAGATATTTTGGTGTAGATAGTAGAGTGGTTCCTGTTTCACTCCCTTAACTAG 816
Db      1837 TTTTGTCTTTTGGTGTAGATAGTAGAGTGGTTCCTGTTTCACTCCCTTAACTAG 1896
OY      817 -CTGACAGTGTCTCCCTGCTTCTGTTCTGTAAGTGTGTAGAAATCACTGCTCGGCT 875
Db      1897 AGCTAGTATAGTGTCTCTTGTCTGTAAGTGTGTAGAAATCACTGCTCGGCT 1956
OY      876 TTACAGTTTAACTATTTAGATATTTCTGAAACATGCTGTCTTGGCAGATACCAAC 935
Db      1957 TTACAGTTTAACTATTTAGATATTTCTGAAACATGCTGTCTTGGCAGATACCAAC 2016
OY      936 ACTGCACTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 995
Db      2017 ACTGCACTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2073
OY      996 T 996
Db      2074 T 2074

```

```

RESULT 5
LOCUS   BU514423
DEFINITION AGNCODR1.10119699 NIH_MGC_134 Mus musculus cDNA clone
IMAGE: 6511045 5', mRNA sequence.
ACCESSION BU514423
VERSION   BU514423.1 GI:22821949
KEYWORDS EST.
SOURCE   Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 978)

```

```

AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE    National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL  Unpublished
COMMENT   Contact: Robert Strusberg, Ph.D.
          Email: cgabbs-remail.nih.gov
          Tissue Procurement: Dr. David Rowe
          cDNA Library Preparation: Invitrogen Corp
          DNA Library Arrayed by: The I.M.A.G.E. Consortium (LNLN)
          DNA sequencing by: Agencourt Bioscience Corporation
          Clone distribution: MGC clone distribution information can be
          found through the I.M.A.G.E. Consortium/LNLN at:
          http://image.llnl.gov
          Plate: LNL4080 row: 0 column: 14
          High quality sequence start: 17
          High quality sequence stop: 714.
FEATURES
    source
        1..978
            /organism="Mus musculus"
            /mol_type="mRNA"
            /db_xref="taxon:10090"
            /clone="IMAGE:6511045"
            /tissue_type="undifferentiated limb"
            /lab_host="DH10B (phage-resistant)"
            /clone_idb="NIH_MGC_134"
            /note="Vector: pCMV-SPORT6.1.cdb; Site.1: Score; Site.2:
            NotI; Cloned unidirectionally. Primer: Oligo dT. Average
            insert size 1.7 kb. Constructed by Reagen, Invitrogen
            Corp. Note: this is a NIH_MGC Library."
BASE COUNT  249 a 224 c 201 g 301 t 3 others
ORIGIN
Query Match      58.4%; Score 594.2; DB 13; Length 978;
Best Local Similarity 85.8%; Pred. No. 1.9e-159;
Matches 762; Conservative 0; Mismatches 94; Indels 32; Gaps 8;
OY      1 CCCCTATTTCTGCTGATTAAGATGCAAAATACCTGTGATGATGATGCTGCTG 60
Db      70 CCCCCTATTTCTGCTGATTAAGATGCAAAATACCTGTGATGATGATGCTGCTG 129
OY      61 CTGAGAACGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 117
Db      130 CTGAGAACGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 189
OY      118 TTTGCGGTGACAGGCTGACAGACCTTCCGCTGAGCTGACAGCTGACAGACGAC 177
Db      190 TTTGCGGTGACAGGCTGACAGACCTTCCGCTGAGCTGACAGCTGACAGACGAC 249
OY      178 TCCCTTAATGACAGCTGACAGCTGACAGCTGATTAATGATGATGATGATGATG 237
Db      250 TCCCTTAATGACAGCTGACAGCTGACAGCTGATTAATGATGATGATGATGATG 309
OY      238 AACCTTGTACGTAGCTGAC---GATGGGTGATGATGATGATGATGATGATGATG 294
Db      310 AACCTTGTACGTAGCTGACGTGATGATGATGATGATGATGATGATGATGATG 369
OY      295 AGTCTCTAAATATGACAGCTGACAGCTGATGATGATGATGATGATGATGATGATG 354
Db      370 AGTCTCTAAATATGACAGCTGATGATGATGATGATGATGATGATGATGATGATG 429
OY      355 TGTGCGCTCTAAGCAAGTATGATGATGATGATGATGATGATGATGATGATGATG 414
Db      430 TGTGCGCTCTAATGATGATGATGATGATGATGATGATGATGATGATGATGATG 487
OY      415 ACAGAGCTGATCAAGAGAGACAGAGATATTTGATGATGATGATGATGATGATG 474
Db      488 ACAGAGCTGATCAAGAGAGACAGAGATATTTGATGATGATGATGATGATGATG 547
OY      475 ACAGTACATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 534
Db      548 ACAGTACATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 607
OY      535 ACGGTTAATTTCTGCAAAAGTATGATGATGATGATGATGATGATGATGATGATG 589

```

Db	ACGGTATATTCTGTCAAAAGTAGTGGAGCAAAATGGCGAATCTTATCAAGTCTCTTGGAG	667
Qy	590 -----TCTTGAGACCCCCCTTCCCCCTGCTGTCCCTCCCAAGATGGGGGCTTGAATCCA	642
Db	668 CATCCCTTGGCGCCGCCCCACCTCTCACAGCGCTCCCTCCCAAGATGGGGCTTGAATCCA	727
Qy	643 TATTAAACTGGCCATCCTCACAAGTGTGTAACCTTAGCAAGTCACTTCTTATAGAACCCCC	702
Db	728 TATTAAAACTGGGCAACCCCTACAGTTCTTAACCTTAGCAAGTCACTTCTTCTTAG--AACCC	785
Qy	703 TTCTTAACGAGCAATATGTCTGACCTGTACATTAAGATCTTTCGTGATATGCA-----T	756
Db	786 TTCTTTACGACAAATATGTCTGACCTGTACATTAAGATCTTTCGTGATTAACCAANAGATTT	845
Qy	757 TCGGAGATTTTTTGTGTAGAT--AGTAGAATGGCGTCTCTGTTTTCACACTTCTTACTC	814
Db	846 TTTGTTTTTTTTTGGTAGCTCAAGTAGAAGTGGGTCTCTGTTTTTGGCCCTACATTACTC	905
Qy	815 --AACTGACTAGTCTCTCCCTTCTGTTTCTTCTAGTAACGTGGGTGTGAAAA	860
Db	906 AGAGCTAGTATAGTGGCTTCTTAGTTTCTTAACCAACTAGGTGGTGGCAA	953

LOCUS	CA318577	826 bp	mRNA	linear	EST 26-NOV-2002
DEFINITION	UI-M-FW0-cbr-o-14-0-UI.r1 NIH_BMAP_FW0		Mus musculus	CDNA clone	
ACCESSION	CA318577				
VERSION	CA318577.1	GI:24536701			
KEYWORDS	EST.				
SOURCE	Mus musculus (house mouse)				
ORGANISM	Mus musculus				
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.				
TITLE	1 (bases 1 to 826)				
AUTHORS	NIH-MGC <a href="http://mgc.nci.nih.gov/">http://mgc.nci.nih.gov/</a> .				
JOURNAL	National Institutes of Health, Mammalian Gene Collection (MGC)				
COMMENT	Unpublished				
	Contact: Robert Strausberg, Ph.D.				
	Email: <a href="mailto:cgapbs-remail.nih.gov">cgapbs-remail.nih.gov</a>				
	Tissue Procurement: Dr. Jim Lin, University of Iowa				
	CDNA Library Preparation: Dr. M. Bento Soares, University of Iowa				
	CDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa				
	DNA Sequencing by: Dr. M. Bento Soares, University of Iowa				
	Clone Distribution: MGC clone distribution information can be				
	found through the I.M.A.G.E. Consortium/LEND at:				
	<a href="http://image.llnl.gov">http://image.llnl.gov</a>				
	This clone was contributed by the Brain Molecular Anatomy Project (BMAP)				
	Seq primer: pyx-5.				

```

FEATURES
SOURCE
    location/Qualifiers
1. 826
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6"
/db_xref="taxon:10090"
/clone="IMAGE: 6813639"
/tissue_type="whole brain"
/dev_stage="embryo 13.5,14.5,16.5,17.5dpc"
/lab_host="DH10B (T1 phage resistant)"
/clone_idb="NH_BMAP_FW0"
/notes="Organ: Brain; Vector: pYX-Asc; Site_1: EcoR I; Site_2: Not I; The library was constructed according Bonaldo, Lennon and Soares, Genome Research, 6:791-806, 1996. Denatured RNA was size fractionated on a 1% agarose gel. First strand cDNA synthesis was primed with oligo-dT primer containing a Not I site. Double strand cDNA was size selected according to mRNA size fraction, ligated with EcoR I adaptor, digested with NotI and then cloned directionally into pYX-Asc vector. The library tag sequence located between the Not I site and the polyA tail is AGCGAGCAG. This library was created for the University

```

BASE COUNT	226	a	195	c	162	g	242	t	1	others
ORIGIN	Iowa Brain Anatomy Project (BMAP): 'Gene Discovery in the Developing Mouse Nervous System', supported by National Institute of Mental Health (NIMH), Hemin Chin, Ph.D., program coordinator."									

Query Match	55.38;	Score 562;	DB 14;	Length 826;
Best Local Similarity	89.78;	Pred. No. 3.1e-150;		
Matches 677; Conservative	0;	Mismatches 56;	Indels 22;	Gaps 6

```
QY      1 CCCCATTCCTGCTCAGATTAAGAATGCCAAAATACCTTGGAACTAAGTCCGTGTG   60
        ||| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db       76 CCCCATTCTGCTCAGATTAAAAATGC AAAATACCTCGAACATAAGTTGGCTGTG  135
```

61 CTGGAACACAGTAAACCACTAAGCGTTTGAGA- --CTTGTCTTAAGAAGACTGCAGC 117  
||||| ||||| | | ||| | ||||| |||||  
Db 136 CTGGAACACACTAAGCACTAAACTCTTTGAGAGACTTCGTTGCTAAGMAGCCGCCAGC 195

118 TTTGGGGCTAAGGGGTGACAGACCCCTCCGCTGCCAGACCGTGATACACACACACAGCC 177  
 ||||| ||||||||||||| | | |||||||||||||||||||||  
 Db 196 TTTGGAGCAGGGGTGACAGACCCCTTCCAGACTTCCAGACCGTGATACACACACAGCC 255

```

07      TCCCTAATGACAGCGTTGCCATGTAAAGCACCCTGAACCTTAACAGCCCATGTCATTACGT   315
Db      TCCCTAATGACAGCGTTGCCATGTAAAGCACCCTGAACCTTAACAGCCCATGTCATTACGT   315

```

Db 316 AACCTTGTACTGTACGTCACAGTGATGAGTGTGACAGCTCAAGCTCTTTGGATTTCATAGTG 375

Db 376 AGTTTCTAAGACGCGGACTTTTCAGACTTTGCAACAGAACTCTGGTTCGTG 435

Db 436 TGTTCCTCATGAAGTATTCCTGTTCTGAGTTGTGGGTGTCTGGGTGGA--GTGTGAA 493

494 ACACGACATGATCAAGGAGACAGACAGTATTTGACTAAATGTGAAGTAGAGATTACTTTT 553

535 ACGGTAATTTCTGTTCAACACAGATGACAAATGCGCGGACGTTTATGACGCTGC-----500

0v	590	-----TCTGTTAGCGCCCGCTCCCGCTGCTCCCTCCCGAGATGGGGCGTTGAGTCCA	642
Db	614	ACGGTTAATTCTGTCACACAGTAGGTGACAAATGGCGGACTTATCAGTGTCTCTTGAG	673

Db	674	CATCTCTTGGGGCCCCCACCCCTCACTGGCGTCCCTCCCAAGATGGGGCGTTGAGTCCA	733
Oy	643	TATTTAACTGGCCATCTCTCAAGTTGCTAACTTAGAAGTCTTTCTTTAGAACCCCC	702

Db	734	TATTTANACTGGCCACCCCTCACAGTT-CTAAGTTAGCAAGTGCTTTCTTTA-GAACCCC	791
Oy	703	TTCTTAACGACGACATATGTCTGACCTGTACTATAA	737

Db 792 TTCTTACGACGCAATATGTCGACACTGATATAA 826

BQ177713					
LOCUS	BQ177713	751 bp	mRNA	linear	EST 30-APR-2000
DEFINITION	UI-M-ERO-dwp-j-03-0-UI.r1 NH_BMP_ERO Mus musculus cDNA clone				

VERSION	B0177713.1	GI:20353205
KEYWORDS	EST.	
SOURCE	Mus musculus (house mouse)	



ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
REFERENCE 1 (bases 1 to 751)  
AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.  
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: [cgapbs-remail.nih.gov](mailto:cgapbs-remail.nih.gov)  
Tissue Procurement: Dr. James Lin, University of Iowa  
CDNA Library preparation: Dr. M. Bento Soares, University of Iowa  
CDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa  
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa  
Clone Distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/ILNI at:  
<http://image.llnl.gov>  
This clone was contributed by the Brain Molecular Anatomy Project  
(BMAP)  
Seq primer: PYX-5.  
Location/Qualifiers  
1. 751  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/strain="C57BL/6"  
/db\_xref="taxon:10090"  
/clone="IMAGE:5700194"  
/tissue\_type="whole brain"  
/dev\_stage="embryo 15.5 dpc"  
/lab\_host="DH10B (T1 phage resistant)"  
/clone\_lib="NIH-BMAP\_ER0"  
/note="Organ: Brain; Vector: pYX-Asc; Site\_1: EcoR I;  
Site\_2: Not I; The library was constructed according to  
Bonaldo, Lennon and Soares, Genome Research, 6:791-806,  
1996. Denatured mRNA was size fractionated on a 1% agarose  
gel. First strand cDNA synthesis was primed with an  
oligo-dT primer containing a Not I site. Double stranded  
cDNA was size selected according to mRNA size fraction,  
ligated with EcoR I adaptor, digested with Not I, and then  
cloned directionally into pYX-Asc vector. The library tag  
sequence located between the Not I site and the polyA tail  
, is GTCGCGGAA. This library was created for the  
University of Iowa Mouse Brain Molecular Anatomy Project  
(BMAP). Gene Discovery in the Developing Mouse Nervous  
System, supported by National Institutes of Mental Health  
(NIMH), Hemin Chin, Ph.D., program coordinator."

BASE COUNT 198 a 180 c 154 g 217 t 2 others  
ORIGIN

Query Match 54.98; Score 558.2; DB 13; Length 751;  
Best Local Similarity 89.6%; Pred. No. 3.7e-149;  
Matches 673; Conservative 0; Mismatches 55; Indels 23; Gaps 6;

QY 1 CCCCTATTCCTGTCGATTAAGAAATGCCAAATACCTTGTAACATAAGTTGGGTGTG 60  
DB 6 CCCCCATTCCTGTCGATTAAGAAATGCCAAATACCTTGTAACATAAGTTGGGTGTG 65  
QY 61 CTGGAACACGTAAGTACTAGTGTGAGAGA--CTTGTCTCTTAAGAGACTGCAGC 117  
DB 66 CTGGAACACGTAAGTACTAGTGTGAGAGA--CTTGTCTCTTAAGAGACTGCAGC 125  
QY 118 TTCTGGGCTCAGGGGCTGACACCTCCCGTAGTCCACAGACCGTGTGACAGACAGCC 177  
DB 126 TTCTGGGCTCAGGGGCTGACACCTCCCGTAGTCCACAGACCGTGTGACAGACAGCC 185  
QY 178 TCCCTAATGACAGCTGCATGTAAGCACTGTAACCTATACCACTGCTCATTTAGCT 237  
DB 186 TCCCTAATGACAGCTGCATGTAAGCACTGTAACCTATACCACTGCTCATTTAGCT 245  
QY 238 AACTTTGACTGTAGCTGAC--GATGGGTGTAACAGCTGCTCTTTGATTTCATAGTG 294  
DB 246 AACTTTGACTGTAGCTGAC--GATGGGTGTAACAGCTGCTCTTTGATTTCATAGTG 305  
QY 295 AGTTCTCTAATAATACCAAGCTGACCGGCTTCTGACAGGCTTTGACAGAACTCTGGCTCTG 354

DB 306 AGTTCTCTAATAATACCAAGCTGACCGGCTTCTGACAGGCTTTGACAGAACTCTGGCTCTG 365  
QY 355 TGTGGCTCTAAGCAAGATATCTGTTCTAGTCGGGTGGTGGCGGAGTGTGAA 414  
DB 366 TGTGGCTCTAAGCAAGATATCTGTTCTAGTCGGGTGGTGGCGGAGTGTGAA 423  
QY 415 ACACGACCTCATCAAGAGAGACAGACAGTATTTGACATCAATATGAGATGATTAATTT 474  
DB 424 ACACGACCTCATCAAGAGAGAGACAGACAGTATTTGACATCAATATGAGATGATTAATTT 483  
QY 475 ACATGACATGTTGACATGAGATTAATCAACATGATTAATCAAGGCTGTAAGCTTTTGA 534  
DB 484 ACATGACATGTTGACATGAGATTAATCAACATGATTAATCAAGGCTGTAAGCTTTTGA 543  
QY 535 ACGGTTATTTCTGCAACAGTATGATGACAAATGGCCGATCTTATCATGTCCTC----- 589  
DB 544 ACGGTTATTTCTGCAACAGTATGATGACAAATGGCCGATCTTATCATGTCCTC----- 603  
QY 590 -----TCTTGAGCCCCCTTCCCTCCCTGCTGCCCTCCAGATGAGGCGTGTGATCA 642  
DB 604 CATCCCTTCGGGCCCCCACCCTCCACCTCCACCTGCTGCTCCAGATGAGGCGTGTGATCA 663  
QY 643 TATTTAACTGGCCATCTCTACACAGTGTGCTAACTAGCAAGTGTCTTTAGACCCCGC 702  
DB 664 TATTTAACTGGCCATCTCTACACAGTGTGCTAACTAGCAAGTGTCTTTAGAC--CCC 720  
QY 703 TTCTTAACGACCAATATGCTGACCTGACT 733  
DB 721 TTCTTAACGACCAATATGCTGACCTGACT 751

RESULT 8  
BU511766  
LOCUS  
DEFINITION AGENCOURT\_10114251 NIH\_MGC\_134 Mus musculus cDNA clone  
BU511766  
VERSION BU511766.1 GI:22817999  
KEYWORDS EST.  
SOURCE Mus musculus (house mouse)  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
REFERENCE 1 (bases 1 to 960)  
AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.  
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: [cgapbs-remail.nih.gov](mailto:cgapbs-remail.nih.gov)  
Tissue Procurement: Dr. David Rowe  
CDNA Library Preparation: Invitrogen Corp  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (ILNI)  
DNA Sequencing by: Agencourt Bioscience Corporation  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/ILNI at:  
<http://image.llnl.gov>  
Plate: L1AM14069 row: e column: 16  
High quality sequence stop: 669.  
Location/Qualifiers  
1. 960  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/db\_xref="taxon:10090"  
/clone="IMAGE:6506583"  
/tissue\_type="undifferentiated limb"  
/lab\_host="DH10B (phage-resistant)"  
/clone\_lib="NIH\_MGC\_134"  
/note="Vector: pCMV-SPORT6.1.cdb; Site\_1: EcoRy; Site\_2:  
NotI; Cloned unidirectionally. Primer: Oligo dr. Average  
insert size 1.7 kb. Constructed by Resgen, Invitrogen  
Corp. Note: this is a NIH\_MGC Library."

BASE COUNT 246 a 218 c 196 g 298 t 2 others





OY	415	ACAGAGCTGATTCMAAAGAGACAGACAGACTATTTTTGACCTAAATGTGAAGTAGAGATTAATTT	474
Dd	361	ACACGAGCTCATCAAAAGAGACAGACAGACTATTTTTTGCATAAATATGAAGTAGAGATTAATTT	420
OY	475	ACACTACTATTTACATGAGTAGAATTCACATGATAATTAAGTGTCCAGGGTTAAAGCTTTTGA	534
Dd	421	ACACTACTATTTACATGAGTAGAATTCACATGATAATTAAGTGTCCAGGGTTAAAGCTTTTGA	480
OY	535	ACGCTTAATTTCTGTCAAACAGTAGATGACAATAATGCCGATCTTAATCAGTGTCTCTTG	594
Dd	481	ACGGGTATTTTCTGTCAAACAGTAGATGACAATAATGCCGATCTTAATCAGTGTCTCTTG	540
OY	595	AGCCCCCCCCTCCCGT 611	
Dd	541	AGCCCCCCCCTCCCGT 557	
RESULT 10			
LOCUS	CAJ19273	793 bp	mRNA linear EST 26-NOV-2007
DEFINITION	UI-M-FW0-cbz-a-24-0-UI-T1 NIH_BMAP_FW0 Mus musculus cDNA clone		
ACCESSION	CAJ19273		
VERSION	CAJ19273.1	GI:24537397	
KEYWORDS	EST.		
SOURCE	Mus musculus (house mouse)		
ORGANISM	Mus musculus		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 793) NIH-MGC http://mgc.nci.nih.gov/. National Institutes of Health, Mammalian Gene Collection (MGC)		
AUTHORS	Unpublished Contact: Robert Strausberg, Ph.D. Email: csqab@remail.nih.gov Tissue Procurement: Dr. Jim Lin, University of Iowa cDNA Library preparation: Dr. M. Bento Soares, University of Iowa DNA library Arrayed by: Dr. M. Bento Soares, University of Iowa DNA Sequencing by: Dr. M. Bento Soares, University of Iowa Clone Distribution: MCC clone distribution information can be found through the I.M.A.G.E. Consortium/LNLN at: http://image.lnl.nh.gov		
TITLE	This clone was contributed by the Brain Molecular Anatomy Project (BMAP)		
JOURNAL			
COMMENT			
FEATURES			
Source	Seq primer: pyx-5. Location/Qualifiers		
	1..793		
	/organism="Mus musculus"		
	/mol_type="mRNA"		
	/strain="C57BL/6"		
	/db_xref="taxon:10090"		
	/cds_image="IMAGE: 6816385"		
	/clone_name="whole brain"		
	/tissue_type="embryo 13.5,14.5,16.5,17.5dpf"		
	/dev_stage="embryo 13.5,14.5,16.5,17.5dpf"		
	/lab_host="DH10B (TI phage resistant)"		
	/clone_lib="NIH BMAP FW0"		
	/note="Organ: Brain; Vector: pYX-Asc; Site_1: EcoR I; Site_2: Not I; The library was constructed according Bonaldo, Lennon and Soares, Genome Research, 6:791-806, 1996. Denatured RNA was size fractionated on a 1% agarose gel. First strand cDNA synthesis was primed with oligo-dT primer containing a Not I site. Double strand cDNA was size selected according to mRNA size fraction, ligated with EcoR I adaptor, digested with NotI and then cloned directionally into pYX-Asc vector. The library tag sequence located between the Not I site and the polyA tail is AGCGAGACAG. This library was created for the University Iowa Brain Anatomy Project (BMAP); 'Gene Discovery in the Developing Mouse Nervous System', supported by National Institute of Mental Health (NIMH), Hemin Chin, Ph.D., program coordinator."		
BASE COUNT	200 a	185 c	165 g
		242 t	1 others

Query Match	52.8%	Score 537.4	DB 14	Length 793
Best Local Similarity	87.2%	Pred. No. 3.7e-143		
Matches 677	Conservative 0	Mismatches 72	Indels 27	Gaps 7
64	AGAACACTAGACNCTAGCTGTGAGAGA---CTTGTCCCTTAAGAAAGACTGCAGCTTC	120		
DB				
1	AGAACACCTAAGCATAACCTCTCTTGGAGAGACCTTCTGTGCTAAGAAAGACCGCAGCTTC	60		
121	TGGGCTAGGGGGTGCAGACCCCTCCGGTAGCTCCAGACCCTGTGACACAGCAGACCTCC	180		
DB				
61	TGGAGCCAGGGGGTGCAGACCCCTTCAGAGTTCCAGACCCTGTGACACAGCAGACCTCC	120		
181	TTAATGACAGCCTGGCAGTATTAACGCACCTGTAACTTATCAGCCCATGCTCATTAAGTAAC	240		
DB				
121	TTAATGACAGCCTGGCAGTATTAACGCACCTGTAACTTATCAGCCCATGCTCATTAAGTAAC	180		
241	TTTGTACTGTAGCTTCAC---GATGGGTATAACAGCTGTGCTCTTGTGATTTCAATAGT	297		
DB				
181	TTTGTACTGTAGCTTCACAGTATGAGTGTGACAGCTGTGATTTCAATAGT	240		
298	TCTCTAAATATACACAGCTGACGGGCTTTCAGAGGCTTTCAGAGACCTGTGCTCCGTGAT	357		
DB				
241	TTTCTAAAGACACAGCAGCTGATAGCTTTTGCAGACTTTAAGACAGACTGTGCTCTGT	300		
358	TGCTCTAAGCAAGTATTTCTGTCTAGCTGTGGGTGTGCTGGGTGAGTGTGTAACA	417		
DB				
301	TGCTCTAAGCAAGTATTTCTGTCTAGCTGTGGGTGTGCTGGGTGAGTGTGTAACA	358		
418	CGAGCTCATCAAGAGCAGACAGTATTTTGTACTAATATGAGTAGAGATTAATTACA	477		
DB				
359	CGACATGATCAAAAGAGCAGACAGTATTTTGTACTAATATGAGTAGAGATTAATTACA	418		
478	CTACATTTAGATGAGATTAATCACTCAATTAAGTGTCAAGGGTAAAGCTTTTAAG	537		
DB				
419	CTACATTTAGATGAGATTAATCACTCAATTAAGTGTCAAGGGTAAAGCTTTTAAG	478		
538	GTTAATTTCTGTCAAAAGAGTATGACAAATGGCCGATCTTATCAGTGTCTC-----	589		
DB				
479	GTTAATTTCTGTCAAAAGAGTATGACAAATGGCCGATCTTATCAGTGTGTGAGCAT	538		
590	----TCTTGAGACCCCTTCCCTCCCTGCTGCTCCCTCCAGATGGGGCTTGAGTGCATAT	645		
DB				
539	CCCTTCCGGGCCCCCACCCTCCACAGCGCCGCTCCCTCCACAGATGGGGCTTGAGTGCATAT	598		
646	TTAAACTGGCCATCTCTCAGTGTGTACTTATAGCAAGTCTTTTATAGAGCCCCCTTC	705		
DB				
599	TTAAACTGGCCATCTCTCAGTGTGTACTTATAGCAAGTCTTTTATAGAGCCCCCTTC	656		
706	TTAAGCAATATGTCTGACCTGTACTATAGATCTTCTGTATTAATCA-----TTCGG	760		
DB				
657	TTAAGCAATATGTCTGACCTGTACTATAGATCTTCTGTATTAATCA-----TTCGG	716		
761	AGATTTTGTGTAGATAGTAGAGTGGCTCTCTTTCACCTCTTACTACTAG	816		
DB				
717	GTTTTTTTTTGTGTAGATAGTAGAGTGGCTCTCTTTCACCTCTTACTACTAG	772		

RESULT 11

B1646446

LOCUS 603227020261 NCI\_CGAP\_Mam3 Mus musculus cDNA clone IMAGE:5319127 5',

DEFINITION mRNA sequence.

ACCESSION B1646446

VERSION B1646446.1 GI:15560682

KEYWORDS EST.

SOURCE Mus musculus (house mouse)

ORGANISM Mus musculus

REFERENCE 1 (bases 1 to 817)

AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.





```

QY      814 CAGCTG--ACIAGTCTCCCTTCGTTTCTAGTAACGGGTGTAGAAATCAGCTGCTGC 871
      |||  - - - - - |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db      726 CAGAACCTGTAGTACCTTCCTAGTGTTCAGCACTAAGTGTGCAAAATCATGTGTGTC 785
QY      872 GGCTTACA--GTTTAAACATTTTATGATTTCTGAAA 909
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db      786 CAGCTTTTACGGGTTTAACTATTTTAAAGAAATCTTTAA 826

RESULT 14
AA817948/c 545 bp mRNA linear EST 03-JUL-1999
LOCUS      UI-R-A0-ag-e-05-0-UI.s1 UI-R-A0 Rattus norvegicus cDNA clone
DEFINITION UI-R-A0-ag-e-05-0-UI 3' similar to dbj|D5274|HUMP02S9 Human
            randomly sequenced mRNA, mRNA sequence.
ACCESSION  AA817948
VERSION     AA817948.1 GI:4198471
KEYWORDS   EST.
SOURCE      Rattus norvegicus (Norway rat)
ORGANISM   Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
            1 (bases 1 to 545)
REFERENCE   1 (bases 1 to 545)
AUTHORS    Bonaldo,M.F., Lennon,G. and Soares,M.B.
TITLE      Normalization and subtraction: two approaches to facilitate gene
            discovery
JOURNAL     Genome Res. 6 (9), 791-806 (1996)
MEDLINE     97044477
PUBMED      8889548
COMMENT     On Feb 17, 1998 this sequence version replaced gi:2887828.
            Contact: Soares, MB
            Coordinated Laboratory for Computational Genomics
            University of Iowa
            375 Newton Road , 4156 MEBRF, Iowa City, IA 52242, USA
            Tel: 319 335 8250
            Fax: 319 335 9565
            Email: bento-soares@uiowa.edu
            The sequence tag present in the cDNA between the NotI site and the
            oligo-dT track served to identify it as a clone from the normalized
            adult kidney library. cDNA library Preparation: M. Fatima Bonaldo,
            Ph.D. Clone distribution: clones will be available through Research
            Genetics This clone is also available through the I.M.A.G.E.
            Consortium at LNL (lnfo@image.lnl.gov). IMAGE ID=1767351
            Seq primer: M13 Forward
            POLYA-No.

FEATURES
    source          location/Qualifiers
    1..545
    /organism="Rattus norvegicus"
    /mol_type="mRNA"
    /strain="Sprague-Dawley"
    /db_xref="taxon:10116"
    /clone="UI-R-A0-ag-e-05-0-UI"
    /dev_stage="adult"
    /lab_host="DH10B (Life Technologies)"
    /clone_lib="UI-R-A0"
    /note="Vector: pTZ19-D-Pac (Pharmacia) with a modified
    polylinker; Site_1: Not I; Site_2: Eco RI; This library
    consists of a mixture of individually tagged normalized
    libraries constructed from rat placenta, adult lung, brain
    liver, kidney, heart, spleen, ovary, and muscle. The tag
    is a string of 3-5 nucleotides present between the Not I
    site and the oligo-dT track which allows identification of
    the library of origin of a clone within the mixture."

BASE COUNT      148 a 120 c 119 g 158 t
ORIGIN
Query Match      51.7%; Score 526; DB 9; Length 545;
Best Local Similarity 100.0%; Pred. No. 5,7e-140;
Matches 526; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY      1 CCCCTATCTTGCTGATTAAGAAATGCCAAATACCTGTGAACTAAGTTCGTTGTC 60

```

```

Db      540 CCCCTATCTTGCTGATTAAGAAATGCCAAATACCTGTGAACTAAGTTCGTTGTC 481
QY      61 CTGAGAACAGCTAAGCACTAAGCTGTGGAGAGACTTGTCCCTTAAGAGACCTCAGCTTC 120
Db      480 CTGAGAACAGCTAAGCACTAAGCTGTGGAGAGACTTGTCCCTTAAGAGACCTCAGCTTC 421
QY      121 TGGGCTCAGGGGTGCAGACCCCTCCGTAAGTCCACAGCGGTGTGACACAGCAGCCTCC 180
Db      420 TGGGCTCAGGGGTGCAGACCCCTCCGTAAGTCCACAGCGGTGTGACACAGCAGCCTCC 361
QY      181 TTAATGACAGCGTCCATGTAAAGCACTTAATATAGCCCATGCTCATTAAGTAAAC 240
Db      360 TTAATGACAGCGTCCATGTAAAGCACTTAATATAGCCCATGCTCATTAAGTAAAC 301
QY      241 TTGTGACTGACGTCAAGATGGGTGTAAACAGCTCTGCTCTTTGATTTCATAGAGATTTC 300
Db      300 TTGTGACTGACGTCAAGATGGGTGTAAACAGCTCTGCTCTTTGATTTCATAGAGATTTC 241
QY      301 CTAAATATACCAAGCTGACACCGGCTCTGCAAGCTTTGAACAGAACTGTGCTGTGTC 360
Db      240 CTAAATATACCAAGCTGACACCGGCTCTGCAAGCTTTGAACAGAACTGTGCTGTGTC 181
QY      361 CTCTAAGCAAGTATTCGTCTCAGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTC 420
Db      180 CTCTAAGCAAGTATTCGTCTCAGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTC 121
QY      421 CGTCATCAAAAGGAGACAGACAGATTTTGTACTAATATGAAGTATTAATTTACACTA 480
Db      120 CGTCATCAAAAGGAGACAGACAGATTTTGTACTAATATGAAGTATTAATTTACACTA 61
QY      481 CATTGTACATGAGTAATTCATGTAATTAAGTGTACAGGGTTAA 526
Db      60 CATTGTACATGAGTAATTCATGTAATTAAGTGTACAGGGTTAA 15

RESULT 15
BQ201343/c 639 bp mRNA linear EST 02-MAY-2002
LOCUS      UI-R-DQ1-clt-n-08-0-UI.s1 UI-R-DQ1 Rattus norvegicus cDNA clone
DEFINITION UI-R-DQ1-clt-n-08-0-UI 3' , mRNA sequence.
ACCESSION  BQ201343
VERSION     BQ201343.1 GI:20417808
KEYWORDS   EST.
SOURCE      Rattus norvegicus (Norway rat)
ORGANISM   Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
            1 (bases 1 to 639)
REFERENCE   1 (bases 1 to 639)
AUTHORS    Bonaldo,M.F., Lennon,G. and Soares,M.B.
TITLE      Normalization and subtraction: two approaches to facilitate gene
            discovery
JOURNAL     Genome Res. 6 (9), 791-806 (1996)
MEDLINE     97044477
PUBMED      8889548
COMMENT     Contact: Soares, MB
            Coordinated Laboratory for Computational Genomics
            University of Iowa
            375 Newton Road , 4156 MEBRF, Iowa City, IA 52242, USA
            Tel: 319 335 8250
            Fax: 319 335 9565
            Email: bento-soares@uiowa.edu
            The sequence contained an oligo-dT track that was present in the
            oligonucleotide that was used to prime the synthesis of first
            strand cDNA and therefore this may represent a bonfide poly A
            tail. The sequence tag present in the cDNA between the NotI site
            and the oligo-dT track served to verify it as a clone from the
            normalized rat cell line R3327-5a library cDNA library Preparation:
            M.B. Soares Lab Clone distribution: clones will be available
            through Research Genetics (www.resgen.com)
            Seq primer: M13 Forward
            POLYA=yes.

```

FEATURES  
sourceLocation/Qualifiers  
1. 639

/organism="Rattus norvegicus"

/mol\_type="mRNA"

/strain="Sprague-Dawley"

/db\_xref="taxon:10116"

/clone="UI-R-DQ1-clt-n-08-0-UI"

/dev\_stage="adult"

/lab\_host="DH10B (Life Technologies)"

/clone\_lib="UI-R-DQ1"

/note="Vector: p773D-Pac (Pharmacia) with a modified polylinker; Site\_1: Not I; Site\_2: Eco RI; The UI-R-DQ1 library is a normalized Rat Cell Line R3327-5A library (nr5a) constructed in p737 PAC vector according to the procedure described by Bonaldo, Lennon & Soares (Normalization and Subtraction: Two Approaches to Facilitate Gene Discovery. Genome Research 6: 791-806, 1996). The oligonucleotide used to prime first strand synthesis contained the sequence tag GGACTAGATC between the Not I cloning site and dt18 stretch. The Rat Cell Line R3327-5A was provided by Mary Hendrix of the University of Iowa.

TAG\_LIB="UI-R-DQ1"  
TAG\_TISSUE="rat cell line R3327-5a"  
TAG\_SEQ="GGACTAGATC"BASE COUNT 183 a 130 c 136 g 190 t  
ORIGIN

Query Match 51.7%; Score 526; DB 13; Length 639;  
Best Local Similarity 100.0%; Pred. No. 6.3e-140;  
Matches 526; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
OY 1 CCCCTATCTTGCCTCAGATTAGAAATGCGAAATACCTGTGACTAAGTGGCTGG 60
DB 541 CCCCTATCTTGCCTCAGATTAGAAATGCGAAATACCTGTGACTAAGTGGCTGG 482
OY 61 CTGAGACACGTAAGCCTAAGCTGTAGAGACTTTGCTTAAAGACATGACGCTTC 120
DB 481 CTGAGACACGTAAGCCTAAGCTGTAGAGACTTTGCTTAAAGACATGACGCTTC 422
OY 121 TGGGCTCAGGGGTGAGACCCCTCCGTAAGTCCGACAGCCGTGTGACACAGCAGCTCC 180
DB 421 TGGGCTCAGGGGTGAGACCCCTCCGTAAGTCCGACAGCCGTGTGACACAGCAGCTCC 362
OY 181 TTAATGACAGCGTGCATGTATAGCAGCCTGTAACTATACAGCCATGCTATTACGTAA 240
DB 361 TTAATGACAGCGTGCATGTATAGCAGCCTGTAACTATACAGCCATGCTATTACGTAA 302
OY 241 TTTGACTGTACGTCAGATGGGGTTAAACAGCTGTCTTTGATTTCATAGTGAATCT 300
DB 301 TTTGACTGTACGTCAGATGGGGTTAAACAGCTGTCTTTGATTTCATAGTGAATCT 242
OY 301 CTAAATATACAGCTACCGGCTTTCGAGGCTTTGAACAGAACTCGGCTCTGTGTGC 360
DB 241 CTAAATATACAGCTACCGGCTTTCGAGGCTTTGAACAGAACTCGGCTCTGTGTGC 182
OY 361 CTCTAACGAAGTATCTGTTCTCTAGTCGTCGGGTGTGTCGGGTGTGGAACAGCA 420
DB 181 CTCTAACGAAGTATCTGTTCTCTAGTCGTCGGGTGTGTCGGGTGTGGAACAGCA 122
OY 421 CGTCATCAAGAGAGACAGACAGTATTTGACTAATATGAAAGTAGAGATTATTTACACTA 480
DB 121 CGTCATCAAGAGAGACAGACAGTATTTGACTAATATGAAAGTAGAGATTATTTACACTA 62
OY 481 CATTTACATGAGATGATTTCAACTGATTTAAAGTGTCAAGGGTAAA 526
DB 61 CATTTACATGAGATGATTTCAACTGATTTAAAGTGTCAAGGGTAAA 16
```

Search completed: August 25, 2003, 23:35:36  
Job time : 2120.56 secs

**THIS PAGE BLANK (USPTO)**



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 25, 2003, 21:58:43 ; Search time 74.6032 Seconds  
(without alignments)  
6016.990 Million cell updates/sec

Title: US-09-717-321A-15  
Perfect score: 1017  
Sequence: 1 ccccatcttgctcagatt.....ccttggtgctgaggttc 1017

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 569978 seqs, 220691566 residues

Total number of hits satisfying chosen parameters: 1139956

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-Processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued\_Patents\_MA: \*  
1: /cgn2\_6/prodata/1/1na/5A.COMB.seq: \*  
2: /cgn2\_6/prodata/1/1na/5B.COMB.seq: \*  
3: /cgn2\_6/prodata/1/1na/5A.COMB.seq: \*  
4: /cgn2\_6/prodata/1/1na/5B.COMB.seq: \*  
5: /cgn2\_6/prodata/1/1na/PCTUS.COMB.seq: \*  
6: /cgn2\_6/prodata/1/1na/Backfills1.seq: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	187.2	18.4	605	3	US-09-385-982-483
C 2	68	6.7	588	3	US-09-385-982-133
C 3	54.2	5.3	7218	1	US-08-232-463-14
C 4	35.4	3.5	2038	1	US-08-181-271A-1
C 5	35.4	3.5	2038	1	US-08-449-315-1
C 6	35.4	3.5	2038	1	US-08-444-803-1
C 7	35.4	3.5	2038	1	US-08-449-043-1
C 8	35.4	3.5	2038	1	US-08-456-265A-1
C 9	35.4	3.5	2038	1	US-08-455-416-1
C 10	35.4	3.5	2038	1	US-08-455-244-1
C 11	35.4	3.5	2038	1	US-08-454-876-1
C 12	35.4	3.5	2038	2	US-08-457-364-1
C 13	35.4	3.5	2038	2	US-08-456-262-1
C 14	35.4	3.5	2038	2	US-08-456-240-1
C 15	35.4	3.5	2038	2	US-08-455-736-1
C 16	35.4	3.5	2038	2	US-08-455-736-1
C 17	35.4	3.5	2038	2	US-08-455-736-1
C 18	35.4	3.5	2038	2	US-08-455-736-1
C 19	35.4	3.5	2038	2	US-08-455-736-1
C 20	35.4	3.5	2038	2	US-08-455-736-1
C 21	35.4	3.5	2038	2	US-08-455-736-1
C 22	35.4	3.5	2038	2	US-08-455-736-1
C 23	35.4	3.5	2038	2	US-08-455-736-1
C 24	35.4	3.5	2038	2	US-08-455-736-1
C 25	35.4	3.5	2038	2	US-08-455-736-1
C 26	35.4	3.5	2038	2	US-08-455-736-1
C 27	35.4	3.5	2038	2	US-08-455-736-1

28	31.6	3.1	4203	2	US-08-866-757-1	Sequence 1, Appl1
29	31.6	3.1	4203	3	US-09-153-593-1	Sequence 1, Appl1
C 30	31.2	3.1	2559	4	US-09-569-098A-109	Sequence 109, Appl
C 31	31.2	3.0	1221	4	US-09-134-001C-365	Sequence 365, Appl
C 32	30.8	3.0	946	4	US-09-599-360B-42	Sequence 42, Appl
C 33	30.6	3.0	684	4	US-09-134-001C-1999	Sequence 1999, Ap
C 34	30.6	3.0	952	3	US-09-174-768-2	Sequence 2, Appl1
C 35	30.6	3.0	58407	4	US-08-916-421B-2	Sequence 2, Appl1
C 36	30.4	3.0	622	3	US-09-109-204-9	Sequence 9, Appl1
C 37	30.4	3.0	632	4	US-09-490-032-9	Sequence 9, Appl1
C 38	30.4	3.0	1422	4	US-09-227-357-18	Sequence 18, Appl1
C 39	30.4	3.0	1975	3	US-09-109-204-4	Sequence 4, Appl1
C 40	30.4	3.0	1975	4	US-09-490-032-4	Sequence 4, Appl1
C 41	30.4	3.0	3010	4	US-09-996-243-313	Sequence 313, App
C 42	30.4	3.0	87350	4	US-08-781-891-79	Sequence 79, Appl
C 43	30.4	3.0	87350	4	US-09-618-166-79	Sequence 79, Appl
C 44	30.4	3.0	87343	4	US-09-791-211-3	Sequence 3, Appl1
C 45	30.2	3.0	13865	3	US-09-009-217-11	Sequence 11, Appl1

ALIGNMENTS

```
RESULT 1
US-09-385-982-483/C
; Sequence 483, Application US/09385982
; Patent No. 6262334
; GENERAL INFORMATION:
; APPLICANT: ENDEGE, WILSON O., ET AL.
; TITLE OF INVENTION: NOVEL HUMAN GENES AND GENE EXPRESSION
; FILE REFERENCE: CCNDA-260XX
; CURRENT APPLICATION NUMBER: US/09/385,982
; EARLIER FILING DATE: 1999-08-30
; EARLIER APPLICATION NUMBER: 09/328,111
; EARLIER FILING DATE: 1999-06-08
; EARLIER APPLICATION NUMBER: 60/117,393
; EARLIER FILING DATE: 1999-01-27
; EARLIER APPLICATION NUMBER: 60/098,639
; EARLIER FILING DATE: 1998-08-31
; NUMBER OF SEQ ID NOS: 344
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 483
; LENGTH: 605
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc.feature
; LOCATION: (1)...(605)
; OTHER INFORMATION: n = A,T,C or G
US-09-385-982-483

Query Match      18.4%   Score 187.2; DB 3; Length 605;
Best Local Similarity 65.5%; Pred. No. 3.2e-52;
Matches 348; Conservative 0; Mismatches 158; Indels 25; Gaps 5;

OY 485 GTACGAGGATATTACGATGATTAAGTCACGCGAAGGTAAGCTTTAAGCGTAATT 544
    ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 597 GGACAGGAGGATTAAGTCGCGGATTAAGCTTTAAGGTAAGCTTTGCAATACGTA 538
    ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 545 TCTGTCAACAGTAG-ATGACAATATGCCGATCTTATGAGTCTCTGTGAGACCCCT 603
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 537 GATTAATAAATAAGTTGTTATTAACCAATATGTTTCTTANGCTTTTCTTTTCTT 478
    ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 604 TCCCTCTGTCTCCCTCCAGATGGGCGTGTGATGATTAATTAATGAGGATCTCA 663
    ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 477 AACACCTGCCATGCTCCCAATGGGCGATTAATTAATTAACGATGCTCTCT 418
    ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 664 CAGTGTCTACTAGACGATGCTTTCTTAAGACCCCTTTAAGCAAGCAATATGCT 723
    ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 417 TATGCTACTAGTATGATGCTTTCTTAAGCAACCCCTTTGAGTGTGATATGCT 358
    ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 724 GACCTGTACTATTAAGATCTTCTGATATATGATTCGAGATTTTGTGATAGTAGTACA 783
```



QY 872 GCGTTACAGTTTAACTATTAGATATCTGAAACATCAGCTCTTGCCAGAGTAC 931  
Db 1363 YY 1422  
QY 932 CAACACTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 970  
Db 1423 YYYYYYYYYYYYGTACCAATCTCTTACTTAACT 1461

## RESULT 4

US-08-181-271A-1/C  
Sequence 1, Application US/08181271A  
Patent No. 5614395  
GENERAL INFORMATION:  
APPLICANT: Ryals, John A.  
APPLICANT: Alexander, Danny C.  
APPLICANT: Beck, James J.  
APPLICANT: Duesing, John H.  
APPLICANT: Friedrich, Leslie B.  
APPLICANT: Goodman, Robert M.  
APPLICANT: Harms, Christian  
APPLICANT: Meins, Jr., Frederick  
APPLICANT: Montoya, Alice  
APPLICANT: Moyer, Mary B.  
APPLICANT: Neuhaus, Jean-Marc  
APPLICANT: Payne, George B.  
APPLICANT: Sperison, Christoph  
APPLICANT: Stinson, Jeffrey R.  
APPLICANT: Utkes, Scott J.  
APPLICANT: Ward, Eric R.  
APPLICANT: Williams, Shericka C.  
TITLE OF INVENTION: CHEMICALLY REGULATABLE AND ANTI-PATHOGENIC  
TITLE OF INVENTION: DNA SEQUENCES AND USES THEREOF  
NUMBER OF SEQUENCES: 106  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: CIBA-GEIGY Corporation  
STREET: 7 Skyline Drive  
CITY: Hawthorne  
STATE: New York  
COUNTRY: USA  
ZIP: 10532  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/181,271A  
FILING DATE: 13-JAN-94  
CLASSIFICATION: 435  
PRIORITY APPLICATION DATA:  
APPLICATION NUMBER: US 08/093,301  
FILING DATE: 16-JUL-1993  
PRIORITY APPLICATION DATA:  
APPLICATION NUMBER: US 07/937,197  
FILING DATE: 6-NOV-1992  
PRIORITY APPLICATION DATA:  
APPLICATION NUMBER: US 07/678,378  
FILING DATE: 1-APR-1991  
PRIORITY APPLICATION DATA:  
APPLICATION NUMBER: US 07/305,566  
FILING DATE: 6-FEB-1989  
PRIORITY APPLICATION DATA:  
APPLICATION NUMBER: US 07/165,667  
FILING DATE: 8-MAR-1988  
PRIORITY APPLICATION DATA:  
APPLICATION NUMBER: US 08/042,847  
FILING DATE: 6-APR-1993  
PRIORITY APPLICATION DATA:  
APPLICATION NUMBER: US 07/632,441  
FILING DATE: 21-DEC-1990  
PRIORITY APPLICATION DATA:

APPLICATION NUMBER: US 07/425,504  
FILING DATE: 20-OCT-1989  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/848,506  
FILING DATE: 6-MAR-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/768,122  
FILING DATE: 27-SEP-1991  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/580,431  
FILING DATE: 7-SEP-1990  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/368,672  
FILING DATE: 20-JUN-1989  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/329,018  
FILING DATE: 24-MAR-1989  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/045,957  
FILING DATE: 12-APR-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Elmer, James Scott  
REGISTRATION NUMBER: 36,129  
REFERENCE/DOCKET NUMBER: S-19825/P1/CGC 1727  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (919)541-8614  
TELEFAX: (919)541-8689  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2038 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 932..1435  
US-08-181-271A-1

## Query Match 3.58; Score 35.4; DB 1; Length 2038;

Best Local Similarity 51.7%; Pred. No. 0.27; Matches 105; Conservative 0; Mismatches 96; Indels 2; Gaps 1;

QY 630 GCGCTGAGTCCATATTTAACTGGCCATCCTCAGATGCTGACTTACCAAGTGGTTT 689  
Db 314 GACGTTAAATACATCAATTTGGACGTAAGAATTTGGATATCTTCACTTTTC 255  
QY 690 CTTTACGACC--CCCTCTTAACGACCAATATGCTGACCTGACTACTAAGATCTTCTG 747  
Db 254 TCTAATTCCTCCCTTAATTTCCAACTTGATGCTGATCATGATCATCATCTTTCTA 195  
QY 748 ATATGCAATTCGAGATTTTGGTAGATAGTAGAAGCGCTTCTCTTTTCACCTTTC 807  
Db 194 TTATTCATTCATCAATCAGCGCTTGATCTGAGATATTAATTAATTTCTATTTCTATTC 135  
QY 808 TTTACTGAGTGCAGTGTGCTTC 830  
Db 134 AATCAGACTGACTTGAATTC 112

## RESULT 5

US-08-449-315-1/C  
Sequence 1, Application US/08449315  
Patent No. 5650505  
GENERAL INFORMATION:  
APPLICANT: Ryals, John A.  
APPLICANT: Alexander, Danny C.  
APPLICANT: Beck, James J.  
APPLICANT: Duesing, John H.  
APPLICANT: Friedrich, Leslie B.  
APPLICANT: Goodman, Robert M.  
APPLICANT: Harms, Christian  
APPLICANT: Meins, Jr., Frederick

APPLICANT: Montoya, Alice  
 APPLICANT: Moyer, Mary B.  
 APPLICANT: Neuhaus, Jean-Marc  
 APPLICANT: Payne, George B.  
 APPLICANT: Spetison, Christoph  
 APPLICANT: Stinson, Jeffrey R.  
 APPLICANT: Uknes, Scott J.  
 APPLICANT: Ward, Eric R.  
 APPLICANT: Williams, Sherlica C.  
 TITLE OF INVENTION: CHEMICALLY REGULATABLE AND ANTI-PATHOGENIC  
 NUMBER OF SEQUENCES: 106  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: CIBA-GEIGY Corporation  
 STREET: 7 Skyline Drive  
 CITY: Hawthorne  
 STATE: New York  
 COUNTRY: USA  
 ZIP: 10532  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: Patent In Release #1.0, Version #1.25  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/449,315  
 FILING DATE: 24-MAY-1995  
 CLASSIFICATION: 800  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: 08/181,271  
 FILING DATE: 13-JAN-94  
 APPLICATION NUMBER: US 08/093,301  
 FILING DATE: 16-JUL-1993  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 07/937,197  
 FILING DATE: 6-NOV-1992  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 07/678,378  
 FILING DATE: 1-APR-1991  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 07/305,566  
 FILING DATE: 6-FEB-1989  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 07/165,667  
 FILING DATE: 8-MAR-1988  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 08/042,847  
 FILING DATE: 6-APR-1993  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 07/632,441  
 FILING DATE: 21-DEC-1990  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 07/425,504  
 FILING DATE: 20-OCT 1989  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 07/848,506  
 FILING DATE: 6-MAR-1992  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 07/768,122  
 FILING DATE: 27-SEP-1991  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 07/580,431  
 FILING DATE: 7-SEP-1990  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 07/368,672  
 FILING DATE: 20-JUN-1989  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 07/329,018  
 FILING DATE: 24-MAR-1989  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 08/045,957  
 FILING DATE: 12-APR-1993  
 ATTORNEY/AGENT INFORMATION:

```

1      NAME: Elmer, James Scott
2      REGISTRATION NUMBER: 36,129
3      REFERENCE/DOCKET NUMBER: S-19825/P1/CGC 1727
4      TELECOMMUNICATION INFORMATION:
5      TELEPHONE: (919)541-8614
6      TELEFAX: (919)541-8689
7      INFORMATION FOR SEQ ID NO: 1:
8      SEQUENCE CHARACTERISTICS:
9      LENGTH: 2038 base pairs
10     TYPE: nucleic acid
11     STRANDEDNESS: single
12     TOPOLOGY: linear
13     MOLECULE TYPE: CDNA
14     FEATURE:
15     NAME/KEY: CDS
16     LOCATION: 932..1435
17     US-08-449-315-1
18
19     Query Match      3.5%; Score 35.4; DB 1; Length 2038;
20     Best Local Similarity 51.7%; Pred. No. 0.27;
21     Matches 105; Conservative 0; Mismatches 96; Indels 2; Gaps 1
22
23 QY      630 GGCCTTAGCTGCATTTTAAATCGCCATCCTCACAGTGTGCTAATCTAGCAAGTCTTT 689
24         | | | | | | | | | | | | | | | | | | | | | | | | | | | |
25 Db      314 GACCTTAAATACATCAATCAATTTTGGACCTAAGAATATTTGGATATCTTCAGTCTTTC 255
26
27 QY      690 CTTTAGAGAC--CCCTTCTTACGAGCAATATGCTCGACTGTACTATAGATCTTTG 747
28         | | | | | | | | | | | | | | | | | | | | | | | | | | | |
29 Db      254 TCTAATTTCCCTCCCTTAATTTCCAACTGTATGTTACTCATGTAATCATCACTTTTCA 195
30
31 QY      748 ATATATGATTCGGAGATTTTTTTTGTAGATAGTGAATGCGTCCCTGTTTACACTTCC 807
32         | | | | | | | | | | | | | | | | | | | | | | | | | | | |
33 Db      194 TTATCTCTCATTCATTCACGCGCTTGATCTGAGATATAATTAATTTCTATTTCATCC 135
34
35 QY      808 TTACTCAGCTGACTAGTGCCTC 830
36         | | | | | | | | | | | |
37 Db      134 AATCAGCACTGACTTGTAGATTC 112
38
39 RESULT 6
40 US-08-444-803-1/c
41 : Sequence 1, Application US/08444803
42 : Patent No. 5654414
43 : GENERAL INFORMATION:
44 : APPLICANT: Ryals, John A.
45 : APPLICANT: Alexander, Danny C.
46 : APPLICANT: Beck, James J.
47 : APPLICANT: Duesing, John H.
48 : APPLICANT: Friedrich, Leslie B.
49 : APPLICANT: Goodman, Robert M.
50 : APPLICANT: Harms, Christian
51 : APPLICANT: Weins, Jr., Frederick
52 : APPLICANT: Montoya, Alice
53 : APPLICANT: Moyer, Mary B.
54 : APPLICANT: Neuhaus, Jean-Marc
55 : APPLICANT: Payne, George B.
56 : APPLICANT: Sperison, Christoph
57 : APPLICANT: Stinson, Jeffrey R.
58 : APPLICANT: Utnes, Scott J.
59 : APPLICANT: Ward, Eric R.
60 : APPLICANT: Williams, Shericca C.
61 : TITLE OF INVENTION: CHEMICALLY REGULATABLE AND ANTI-PATHOGENIC
62 : TITLE OF INVENTION: DNA SEQUENCES AND USES THEREOF
63 : NUMBER OF SEQUENCES: 106
64 : CORRESPONDENCE ADDRESS:
65 : ADDRESSEE: CIBA-GEIGY Corporation
66 : STREET: 7 Skyline Drive
67 : CITY: Hawthorne
68 : STATE: New York
69 : COUNTRY: USA
70 : ZIP: 10532
71 : COMPUTER READABLE FORM:
72 : MEDIUM TYPE: Floppy disk

```

```

1 COMPUTER: IBM PC COMPATIBLE
2 OPERATING SYSTEM: PC-DOS/MS-DOS
3 SOFTWARE: Patentin Release #1.0, Version #1.25
4
5 CURRENT APPLICATION DATA:
6 APPLICATION NUMBER: US/08/444,803
7 FILING DATE: 19-MAY-1995
8 CLASSIFICATION: 536
9
10 PRIOR APPLICATION DATA:
11 APPLICATION NUMBER: 08/181,271
12 FILING DATE: 13-JAN-94
13 APPLICATION NUMBER: US 08/093,301
14 FILING DATE: 16-JUL-1993
15
16 PRIOR APPLICATION DATA:
17 APPLICATION NUMBER: US 07/937,197
18 FILING DATE: 6-NOV-1992
19
20 PRIOR APPLICATION DATA:
21 APPLICATION NUMBER: US 07/678,378
22 FILING DATE: 1-APR-1991
23
24 PRIOR APPLICATION DATA:
25 APPLICATION NUMBER: US 07/305,566
26 FILING DATE: 6-FEB-1989
27
28 PRIOR APPLICATION DATA:
29 APPLICATION NUMBER: US 07/165,667
30 FILING DATE: 8-MAR-1988
31
32 PRIOR APPLICATION DATA:
33 APPLICATION NUMBER: US 08/042,847
34 FILING DATE: 6-APR-1993
35
36 PRIOR APPLICATION DATA:
37 APPLICATION NUMBER: US 07/632,441
38 FILING DATE: 21-DEC-1990
39
40 PRIOR APPLICATION DATA:
41 APPLICATION NUMBER: US 07/425,504
42 FILING DATE: 20-OCT-1989
43
44 PRIOR APPLICATION DATA:
45 APPLICATION NUMBER: US 07/848,506
46 FILING DATE: 6-MAR-1992
47
48 PRIOR APPLICATION DATA:
49 APPLICATION NUMBER: US 07/768,122
50 FILING DATE: 27-SEP-1991
51
52 PRIOR APPLICATION DATA:
53 APPLICATION NUMBER: US 07/580,431
54 FILING DATE: 7-SEP-1990
55
56 PRIOR APPLICATION DATA:
57 APPLICATION NUMBER: US 07/368,672
58 FILING DATE: 20-JUN-1989
59
60 PRIOR APPLICATION DATA:
61 APPLICATION NUMBER: US 07/329,018
62 FILING DATE: 24-MAR-1989
63
64 PRIOR APPLICATION DATA:
65 APPLICATION NUMBER: US 08/045,957
66 FILING DATE: 12-APR-1993
67
68 ATTORNEY/AGENT INFORMATION:
69 NAME: Elmer, James Scott
70 REGISTRATION NUMBER: 36,129
71 REFERENCE/DOCKET NUMBER: S-19825/P1/GGC 1727
72 TELECOMMUNICATION INFORMATION:
73 TELEPHONE: (919)541-8614
74 TELEFAX: (919)541-8689
75 INFORMATION FOR SEQ ID NO: 1:
76 SEQUENCE CHARACTERISTICS:
77 LENGTH: 2038 base pairs
78 TYPE: nucleic acid
79 STRANDEDNESS: single
80 TOPOLOGY: linear
81 MOLECULE TYPE: cDNA
82 FEATURE:
83 NAME/KEY: CDS
84 LOCATION: 932..1435
85
86 US-08-444-803-1

```

Query Match	3.5%	Score 35.4	DB 1	Length 2038
Best Local Similarity	51.7%	Pred No. 0.27		
Matches 105	Conservative 0	Mismatches 96	Indels 2	Gaps 1

QY 630 GGGGTGAGACCAATTATTAACCTGGCAATCCCAAGTGTCACTTAGCAAGGCTTT 689

QY 314 GACGTTAAATACATCAATCAAAATTTGGACGCTAAAGAAATTTGGATATTCCTTCAGTCTTTC 255

QY 690 CTTTAGGACC - CCCTTCTTACGAGCATATGTCGACCTGTACTATAGACTCTTCTG 747

Db 254 TCTAAATTCCTCCCTCTAATTTCCAACTGATGTACTCATGTACTCATCAATCTTCTA 195

QY 748 ATATATGCATTCGGAGATTTTTTGGTAGATAGTAAATAGCGCTTCGTTTACCTTCC 807

Db 194 TTATGTCATTCATCAATCAGCGCTTGATCTGGAGATAAATTAATTTCTATTTCTTATCC 135

QY 808 TTACTCAGCTACTAGTCTTC 830

Db 134 AATCAGACACTGACTTATGATTC 112

RESULT 7  
US-08-449-043-1/c  
: Sequence 1, Application US/08449043  
: Patent No. 5689044  
: GENERAL INFORMATION:  
: APPLICANT: Ryals, John A.  
: APPLICANT: Alexander, Danny C.  
: APPLICANT: Beck, James J.  
: APPLICANT: Duesting, John H.  
: APPLICANT: Friedrich, Leslie B.  
: APPLICANT: Goodman, Robert M.  
: APPLICANT: Harms, Christian  
: APPLICANT: Meins, Jr., Frederick  
: APPLICANT: Montoya, Alice  
: APPLICANT: Moyer, Mary B.  
: APPLICANT: Neuhaus, Jean-Marc  
: APPLICANT: Payne, George B.  
: APPLICANT: Sperison, Christoph  
: APPLICANT: Stinson, Jeffrey R.  
: APPLICANT: Uknes, Scott J.  
: APPLICANT: Ward, Eric R.  
: APPLICANT: Williams, Shericca C.  
: TITLE OF INVENTION: CHEMICALLY REGULATABLE AND ANTI-PATHOGENIC  
: TITLE OF INVENTION: DNA SEQUENCES AND USES THEREOF  
: NUMBER OF SEQUENCES: 106  
: CORRESPONDENCE ADDRESS:  
: ADDRESSEE: CIBA-GEIGY Corporation  
: STREET: 7 SKYLINE Drive  
: City: Hawthorne  
: STATE: New York  
: COUNTRY: USA  
: ZIP: 10532  
: COMPUTER READABLE FORM:  
: MEDIUM TYPE: Floppy disk  
: COMPUTER: IBM PC compatible  
: OPERATING SYSTEM: PC-DOS/MS-DOS  
: SOFTWARE: Patentin Release #1.0, Version #1.25  
: CURRENT APPLICATION DATA:  
: APPLICATION NUMBER: US/08/449,043  
: FILING DATE: 24-MAY-1995  
: CLASSIFICATION: 800  
: PRIOR APPLICATION DATA:  
: APPLICATION NUMBER: 08/181,271  
: FILING DATE: 13-JAN-94  
: APPLICATION NUMBER: US 08/093,301  
: FILING DATE: 16-JUL-1993  
: PRIOR APPLICATION DATA:  
: APPLICATION NUMBER: US 07/937,197  
: FILING DATE: 6-NOV-1992  
: PRIOR APPLICATION DATA:  
: APPLICATION NUMBER: US 07/678,378  
: FILING DATE: 1-APR-1991  
: PRIOR APPLICATION DATA:  
: APPLICATION NUMBER: US 07/305,566  
: FILING DATE: 6-FEB-1989







Db 314 GACCTTAATACTATATAATTTGGACGTAAGAAATTTGGATATTTCTCAGTTCTTTC 255  
QY 690 CTTAGAGACC--CCCTCTTAACGAGCAATATGTCGACCTGTAATAATCTTCTG 747  
Db 254 TCTAATTTCCCTTCCTTAATTTCCAACTGTGATGTACGATGCTATCACTTTCTA 195  
QY 748 ATAAATGATTCGGAGATTTTGTGTAGATAGTAAGTGGCTTCGTTTTCACCTTC 807  
Db 194 TTATCTGATTCGATCAACGCGGTGATGTGAGTAAATAATTTCTATTTCTATCCC 135  
QY 808 TTTACTCAGCTGACTAGTGTTC 830  
Db 134 AATCAGACTTGACTTTAGATTTC 112

RESULT 10  
US-08-455-244-1/c  
; Sequence 1, Application US/08455244  
; Patent No. 5789214  
; GENERAL INFORMATION:  
; APPLICANT: Ryals, John A.  
; APPLICANT: Alexander, Danny C.  
; APPLICANT: Beck, James J.  
; APPLICANT: Duesing, John H.  
; APPLICANT: Friedrich, Leslie B.  
; APPLICANT: Goodman, Robert M.  
; APPLICANT: Harms, Christian  
; APPLICANT: Helms, Jr., Frederick  
; APPLICANT: Montoya, Alice  
; APPLICANT: Moyer, Mary B.  
; APPLICANT: Neuhaus, Jean-Marc  
; APPLICANT: Payne, George B.  
; APPLICANT: Sperlison, Christoph  
; APPLICANT: Stinson, Jeffrey R.  
; APPLICANT: Uknes, Scott J.  
; APPLICANT: Ward, Eric R.  
; APPLICANT: Williams, Sherlisa C.  
; TITLE OF INVENTION: CHEMICALLY REGULATABLE AND ANTI-PATHOGENIC  
; TITLE OF INVENTION: DNA SEQUENCES AND USES THEREOF  
; NUMBER OF SEQUENCES: 106  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: CIBA-GEIGY Corporation  
; STREET: 7 Skyline Drive  
; CITY: Hawthorne  
; STATE: New York  
; COUNTRY: USA  
; ZIP: 10532  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; OPERATING SYSTEM: IBM PC compatible  
; SOFTWARE: Patentin Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/455,244  
; FILING DATE: 31-MAY-1995  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/181,271  
; FILING DATE: 13-JAN-94  
; APPLICATION NUMBER: US 08/093,301  
; FILING DATE: 16-JUL-1993  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/937,197  
; FILING DATE: 6-NOV-1992  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/678,378  
; FILING DATE: 1-APR-1991  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/305,566  
; FILING DATE: 6-FEB-1989  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/165,667

; FILING DATE: 8-MAR-1988  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/042,847  
; FILING DATE: 6-APR-1993  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/632,441  
; FILING DATE: 21-DEC-1990  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/425,504  
; FILING DATE: 20-OCT-1989  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/848,506  
; FILING DATE: 6-MAR-1992  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/768,122  
; FILING DATE: 27-SEP-1991  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/580,431  
; FILING DATE: 7-SEP-1990  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/368,672  
; FILING DATE: 20-JUN-1989  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/329,018  
; FILING DATE: 24-MAR-1989  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/045,957  
; FILING DATE: 12-APR-1993  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Elmer, James Scott  
; REGISTRATION NUMBER: 36,129  
; REFERENCE/DOCKET NUMBER: S-19825/P1/CGC 1727  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (919)541-8614  
; TELEFAX: (919)541-8689  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 2038 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: cDNA  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 932..1435  
; US-08-455-244-1

Query Match 3.5%; Score 35.4; DB 1; Length 2038;  
Best Local Similarity 51.7%; Pred. No. 0.27; Mismatches 96; Indels 2; Gaps 1;  
Matches 105; Conservative 0;

QY 630 GCGGTGAGTCATATTTAACTGCCATCTCACAGTTGCTAATTAGCAAGTCTTT 689  
Db 314 GACCTTAATACTATATAATTTGGACGTAAGAAATTTGGATATTTCTCAGTTCTTTC 255  
QY 690 CTTAGAGACC--CCCTCTTAACGAGCAATATGTCGACCTGTAATAATCTTCTG 747  
Db 254 TCTAATTTCCCTTCCTTAATTTCCAACTGTGATGTACGATGCTATCACTTTCTA 195  
QY 748 ATAAATGATTCGGAGATTTTGTGTAGATAGTAAGTGGCTTCGTTTTCACCTTC 807  
Db 194 TTATCTGATTCGATCAACGCGGTGATGTGAGTAAATAATTTCTATTTCTATCCC 135  
QY 808 TTTACTCAGCTGACTAGTGTTC 830  
Db 134 AATCAGACTTGACTTTAGATTTC 112

RESULT 11  
US-08-454-876-1/c  
; Sequence 1, Application US/08454876  
; Patent No. 5804693  
; GENERAL INFORMATION:

```

      FILING DATE: 20-JUN-1989
      PRIOR APPLICATION DATA:
      APPLICATION NUMBER: US 07/329,018
      FILING DATE: 24-MAR-1989
      PRIOR APPLICATION DATA:
      APPLICATION NUMBER: US 08/045,957
      FILING DATE: 12-APR-1993
      ATTORNEY/AGENT INFORMATION:
      NAME: Elmer, James Scott
      REGISTRATION NUMBER: 36,129
      REFERENCE/DOCKET NUMBER: S-19825/P1/GC 1727
      TELECOMMUNICATION INFORMATION:
      TELEPHONE: (919)541-8614
      TELEFAX: (919)541-8689
      INFORMATION FOR SEQ ID NO: 1:
      SEQUENCE CHARACTERISTICS:
      LENGTH: 2038 base pairs
      TYPE: nucleic acid
      STRANDEDNESS: single
      TOPOLOGY: linear
      MOLECULE TYPE: CDNA
      FEATURE:
      NAME/KEY: CDS
      LOCATION: 932..1435
      US-08-454-876-1

Query Match          3.5%; Score 35.4; DB 1; Length 2038;
Best Local Similarity 51.7%; Pred. No. 0.27; Mismatches 96; Indels 2; Gaps 1
Matches 105; Conservative 0;

QY 630 GGCCTGAGTCATATTTAAATCGCCATCCCTCACAGTTGCTAACTTAGCAAGTCTTTT 689
   ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 314 GACGTGTAATTAATACATCAAAATTTGGAGCTAAGAATATTTGGATTTCTGAGTCTTTC 255
QY 690 CTTTAGAGACC--CCCTTCTTAACGACGAAATGTCTGACCGTACTATAAGATCTTTTG 747
   ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 254 TCTAATTTCTCTCCCTTAATTTCCAACTTGATGTACTCAAGTACTAATCAATACCTTTCTA 195
QY 748 AATAATGCAATCGAGATATTTTGTGTAGATAGTAGAAGTGGCTTCTGTTTACACCTTCC 807
   ||||| ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 194 TTATCTCATTTCCAAATCAACGGCTTGATCTGAGAGTAAATTAATTTCTATTTCTTATGCC 135
QY 808 TTACTACAGCTGACTAGTGTCTC 830
   | | ||||| |||||
Db 134 AATCAGCACTTGACTTAGATTTC 112

RESULT 12
US-08-457-364-1/C
: Sequence 1, Application US/08457364
: Patent No. 5847258
: GENERAL INFORMATION:
: APPLICANT: Ryals, John A.
: APPLICANT: Alexander, Danny C.
: APPLICANT: Beck, James J.
: APPLICANT: Duesing, John H.
: APPLICANT: Friedrich, Leslie B.
: APPLICANT: Goodman, Robert M.
: APPLICANT: Harms, Christian
: APPLICANT: Melns, Jr., Frederick
: APPLICANT: Montoya, Alice
: APPLICANT: Moyer, Mary B.
: APPLICANT: Neunhaus, Jean-Marc
: APPLICANT: Payne, George B.
: APPLICANT: Sperison, Christoph
: APPLICANT: Stinson, Jeffrey R.
: APPLICANT: Ukens, Scott J.
: APPLICANT: Ward, Eric R.
: APPLICANT: Williams, Shericea C.
: TITLE OF INVENTION: CHEMICALLY REGULATABLE AND ANTI-PATHOGENIC
: NUMBER OF SEQUENCES: 106
: CORRESPONDENCE ADDRESS: 106

```









**THIS PAGE BLANK (USPTO)**



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 25, 2003, 22:04:33 ; Search time 248.548 seconds  
(without alignments)  
9198.941 Million cell updates/sec

Title: US-09-717-321A-15

Perfect score: 1017

Sequence: 1 cccctattctgtctcagatt.....ccttgggtctgtgaggttc 1017

Scoring table:

IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 1517243 seqs, 1124081882 residues

Total number of hits satisfying chosen parameters: 3034486

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Published Applications\_NA:\*

- 1: /cgn2\_6/ptodata/2/pubpna/US07\_PUBCOMB.seq:\*
- 2: /cgn2\_6/ptodata/2/pubpna/PCR\_NEW\_PUB.seq:\*
- 3: /cgn2\_6/ptodata/2/pubpna/US06\_NEW\_PUB.seq:\*
- 4: /cgn2\_6/ptodata/2/pubpna/US06\_PUBCOMB.seq:\*
- 5: /cgn2\_6/ptodata/2/pubpna/US07\_NEW\_PUB.seq:\*
- 6: /cgn2\_6/ptodata/2/pubpna/US07\_PUBCOMB.seq:\*
- 7: /cgn2\_6/ptodata/2/pubpna/US08\_NEW\_PUB.seq:\*
- 8: /cgn2\_6/ptodata/2/pubpna/US08\_PUBCOMB.seq:\*
- 9: /cgn2\_6/ptodata/2/pubpna/US09A\_PUBCOMB.seq:\*
- 10: /cgn2\_6/ptodata/2/pubpna/US09B\_PUBCOMB.seq:\*
- 11: /cgn2\_6/ptodata/2/pubpna/US09C\_PUBCOMB.seq:\*
- 12: /cgn2\_6/ptodata/2/pubpna/US09\_NEW\_PUB.seq:\*
- 13: /cgn2\_6/ptodata/2/pubpna/US10A\_PUBCOMB.seq:\*
- 14: /cgn2\_6/ptodata/2/pubpna/US10B\_PUBCOMB.seq:\*
- 15: /cgn2\_6/ptodata/2/pubpna/US10\_NEW\_PUB.seq:\*
- 16: /cgn2\_6/ptodata/2/pubpna/US60\_NEW\_PUB.seq:\*
- 17: /cgn2\_6/ptodata/2/pubpna/US60\_PUBCOMB.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	290.6	28.6	1644	US-10-153-668-48	Sequence 48, Appl
2	290.6	28.6	1646	US-10-153-668-46	Sequence 46, Appl
3	290.6	28.6	2051	US-09-923-302-255	Sequence 255, App
4	290.6	28.6	2941	US-10-198-846-9874	Sequence 9874, Ap
5	285.2	28.0	1232	US-09-880-107-1632	Sequence 1632, Ap
6	283.2	27.8	2313	US-10-198-846-12919	Sequence 12919, A
7	195.2	19.0	594	US-10-066-543-2864	Sequence 2864, Ap
8	193.2	19.0	602	US-09-878-178-1654	Sequence 1654, Ap
9	193.2	19.0	602	US-10-046-935-1654	Sequence 1654, Ap
10	193.2	19.0	602	US-10-146-502-1654	Sequence 1654, Ap
11	187.8	18.5	567	US-10-066-543-2878	Sequence 2878, Ap
12	187.8	18.5	567	US-10-066-543-2878	Sequence 2878, Ap
13	187.2	18.4	605	US-09-871-161-483	Sequence 483, App
14	185.4	18.2	477	US-10-066-543-3022	Sequence 3022, App
15	172.2	16.9	432	US-09-960-352-7612	Sequence 7612, App
16	164	16.1	2167	US-10-101-510-500	Sequence 500, App

17	147.4	14.5	409	US-10-066-543-13	Sequence 13, Appl
18	126.4	12.4	285	US-09-294-093B-506	Sequence 506, App
19	123.2	12.1	416	US-09-960-352-8781	Sequence 8781, Ap
20	108.8	10.7	293	US-10-066-543-3077	Sequence 3077, Ap
21	103.2	10.1	3740	US-09-764-891-9984	Sequence 9984, Ap
22	98.2	9.7	581	US-09-998-598-972	Sequence 972, App
23	91.6	9.0	565	US-09-998-598-966	Sequence 296, App
24	85.6	8.4	492	US-09-878-178-1503	Sequence 1503, App
25	85.6	8.4	492	US-10-046-935-1503	Sequence 1503, Ap
26	85.6	8.4	492	US-10-046-935-1503	Sequence 1503, Ap
27	82.2	8.1	398	US-09-918-995-34509	Sequence 34509, A
28	68	6.7	486	US-09-871-161-133	Sequence 133, App
29	62.2	6.1	426	US-09-960-352-7342	Sequence 7342, Ap
30	59	5.8	404	US-09-960-352-14206	Sequence 14206, A
31	57.6	5.7	1651	US-10-198-846-13019	Sequence 13019, A
32	52.4	5.2	162	US-09-920-100A-1549	Sequence 1549, Ap
33	52.4	5.2	162	US-10-033-528-1549	Sequence 1549, Ap
34	40.8	4.0	269	US-09-923-876-6297	Sequence 6297, Ap
35	37.4	3.7	565	US-10-311-455-1504	Sequence 1504, Ap
36	37.2	3.7	565	US-10-027-632-133365	Sequence 133365, Ap
37	37	3.6	260209	US-10-025-966A-23	Sequence 23, Appl
38	37	3.6	260209	US-10-265-071-23	Sequence 23, Appl
39	36.6	3.6	1581	US-10-198-846-6480	Sequence 6480, Ap
40	36.4	3.6	725	US-10-027-632-11287	Sequence 11287, A
41	36.2	3.6	968	US-10-027-632-121396	Sequence 121396, A
42	36	3.5	5098	US-10-192-867-1	Sequence 1, Appl1
43	35.2	3.5	560	US-10-027-632-268476	Sequence 268476, Ap
44	34.8	3.4	454	US-10-027-632-182180	Sequence 182180, Ap
45	34.8	3.4	454	US-10-027-632-182181	Sequence 182181, Ap

#### ALIGNMENTS

RESULT 1  
US-10-153-668-48  
Sequence 48, Application US/10153668  
Publication No. US20030092616A1  
GENERAL INFORMATION:  
APPLICANT: HONDA, Goichi  
APPLICANT: MATSUDA, AKIO  
APPLICANT: MORIMATSU, Shuji  
APPLICANT: ISHIZAWA, Kenya  
TITLE OF INVENTION: STAT6 Activating Gene  
FILE REFERENCE: 1254-0207P  
CURRENT APPLICATION NUMBER: US/10/153,668  
CURRENT FILING DATE: 2002-05-24  
PRIOR APPLICATION NUMBER: US 60/293,172  
PRIOR FILING DATE: 2001-05-25  
PRIOR APPLICATION NUMBER: US 60/316,031  
PRIOR FILING DATE: 2001-08-31  
PRIOR APPLICATION NUMBER: US 60/328,403  
PRIOR FILING DATE: 2001-10-12  
PRIOR APPLICATION NUMBER: JP 2001-157043  
PRIOR FILING DATE: 2001-05-25  
PRIOR APPLICATION NUMBER: JP 2001-260681  
PRIOR FILING DATE: 2001-08-30  
PRIOR APPLICATION NUMBER: JP 2001-313175  
PRIOR FILING DATE: 2001-10-10  
NUMBER OF SEQ ID NOS: 488  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 48  
LENGTH: 1644  
TYPE: DNA  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: CDS  
LOCATION: (109)..(447)  
US-10-153-668-48  
Query Match 28.6%; Score 290.6; DB 14; Length 1644;  
Best Local Similarity 64.5%; Pred. No. 2.5e-83;  
Matches 664; Conservative 0; Mismatches 304; Indels 61; Gaps 13;



OY	469	TAATTTACACATGATTTGACATGGAGTAA---TTCACTGAAATTAAGATGACAGGGTAA	525
Db	869	--ATTTACACATGATTTGACAGAGATGAAGATGTACAGGGTAAAGAACTCTAAAGAGTTA	926
OY	526	AGCTTTTAAACGGTAAATTTCTGTCAACACAGTACATGACAAATGGCGGATCTTATCAGTG	585
Db	927	ATTTCTGTCAAAATGACAGATGATGATGAAGAAAGAGTGGTATTTATACAGGAATGTTTTCT	986
OY	586	TCCTCTTTGAGCCCCCTTCCCGCTGCTGTGCCCTGCCAGATGGGGGCTTGAATCCATAT	645
Db	987	TAACTTTTCTCTTCTTCTTACACCTGCGCATGCTGCCAAATGGGCAATTTAATTCATCT	1048
OY	646	TTAAACATGGCCATCCTCACAGTTGCTTAACCTTAGCAAGTGCCTTTCTTTAGGACCCCTTC	705
Db	1047	TTAAACATGGTGTCTGTCTGTAGTCCGTAACTTAAGTAAAGTGCTTTCTTTATAGAACCCCTTC	1106
OY	706	TTAAACGAGCAATATGTCTGACACTGTACATATAAGATCTTTCGATATATGATTCGGAAAT	765
Db	1107	TGACTGTAGCAATATAGCCCT-CCTTGTATATATATAAATCTTTCGATATATGATTCAGAAAGTT	1165
OY	766	TTTTTGGATAGTATAGATAGAGTGCCTGCTGTTTCACCTTCTCTTACTACAGCTGACTAGT	825
Db	1166	TTTTTGTGCGATTATGATTAAGATGCTTTCATGTTAC-----TTATATGAGACTAATATAGT	1220
OY	826	GCTTCCCTTCCTTTCTTTCTAGTAACTGCGGTGTAGAAATCAGCGTCCGCGCTTACAGTTT	885
Db	1221	GCTTCCCTTAATTTTCTTCTGTAGTAACTGAGGTAAATCATGTGTGACACCTTATATAGTTT	1280
OY	886	TAAACTATTTAGATA-----TTCTGAAACATCAGCTGTCTGCCAGAG	928
Db	1281	TAAATATTTAGATATAATTTCTTAAACTATGAACTCTTAAACATCAGCTGTCTGCCAGAT	1340
OY	929	TACCACACATGTCATGTGATGTGATGCCGCCCTCTAGACCTCACCACGCGGACACATG	988
Db	1341	TACCGACAGCTGTCACTGTACCAATACGT-ACCCTCTTACCTCGCGCCACGGGAGACACG	1399
OY	989	CTTCCGGTA 997	
Db	1400	CTCTCTGTGA 1408	

Query Match	Best Local Similarity	Matches	Conservative	Score	DB %	Length	2051;
28.6%	64.5%	0	0	290.6	DB 9	1686	
664	664	0	0	290.6	DB 9	1686	
1	1	0	0	290.6	DB 9	1686	
755	755	0	0	290.6	DB 9	1686	
58	58	0	0	290.6	DB 9	1686	
815	815	0	0	290.6	DB 9	1686	
118	118	0	0	290.6	DB 9	1686	
875	875	0	0	290.6	DB 9	1686	
170	170	0	0	290.6	DB 9	1686	
935	935	0	0	290.6	DB 9	1686	
230	230	0	0	290.6	DB 9	1686	
995	995	0	0	290.6	DB 9	1686	
290	290	0	0	290.6	DB 9	1686	
1055	1055	0	0	290.6	DB 9	1686	
350	350	0	0	290.6	DB 9	1686	
1113	1113	0	0	290.6	DB 9	1686	
410	410	0	0	290.6	DB 9	1686	
1167	1167	0	0	290.6	DB 9	1686	
469	469	0	0	290.6	DB 9	1686	
1215	1215	0	0	290.6	DB 9	1686	
526	526	0	0	290.6	DB 9	1686	
1273	1273	0	0	290.6	DB 9	1686	
586	586	0	0	290.6	DB 9	1686	
1333	1333	0	0	290.6	DB 9	1686	
646	646	0	0	290.6	DB 9	1686	
1393	1393	0	0	290.6	DB 9	1686	
706	706	0	0	290.6	DB 9	1686	
1453	1453	0	0	290.6	DB 9	1686	
766	766	0	0	290.6	DB 9	1686	
1512	1512	0	0	290.6	DB 9	1686	
826	826	0	0	290.6	DB 9	1686	
1567	1567	0	0	290.6	DB 9	1686	
886	886	0	0	290.6	DB 9	1686	
1627	1627	0	0	290.6	DB 9	1686	



[illegible]

```

: Publication No. US2003009974A1
: GENERAL INFORMATION:
: APPLICANT: Lillie, James
: APPLICANT: Xu, Yongzuo
: APPLICANT: Wang, Youzhen
: APPLICANT: Steinmann, Kathleen
: TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS
: TITLE OF INVENTION: FOR IDENTIFICATION, ASSESSMENT, PREVENTION, AND
: TITLE OF INVENTION: THERAPY OF BREAST CANCER
: FILE REFERENCE: MRI-049
: CURRENT APPLICATION NUMBER: US/10/198,846
: CURRENT FILING DATE: 2002-07-18
: PRIOR APPLICATION NUMBER: 60/306,220
: PRIOR FILING DATE: 2001-07-18
: NUMBER OF SEQ ID NOS: 14084
: SOFTWARE: FASTSEQ for Windows Version 4.0
: SEQ ID NO 12919
: LENGTH: 2313
: TYPE: DNA
: ORGANISM: Homo sapiens
: US-10-198-846-12919

```

Oy	709	ACGGGCAATATGTCGACCTGCTACTATAAGCTCTTCTGAAATGACATTGGGCAATTTT	768
Db	1923	CTGAGCAATATGCTT-CCTGTATATATAAAATCTTTCTGATATGCAATTGAAAGTTTTT	1981
Oy	769	TTGGTAGATAGTAGAAGTGGCGTTTCCTGTTTTCACCTTCCTTTACTCAGCTAGTAGTCT	828
Db	1982	TTTGTCGATTTAGTAAAGTGGCTTTTCCATGTTTAC-----TTTATTCAGAGCTTAATTAAGTCT	2036
Oy	829	TCCCTTGATGTTTCTAGTATACAGGGGTGTAGAAATCAGCTGGCGGGGCTTACAGTTTAA	888
Db	2037	TTCCTTAGTTTCTTCTAGTATACAGTGTAAAAATATGTGTGGCAGCTTTTATAGTTTAA	2096
Oy	889	ACATATTTAGATA-----TCTGAAGACATCACTGTCTTGCCAGAGTAC	931
Db	2097	AATATTTTAGTAAATTTCTTAAACTATGAAACCTTCTTAAACATCACTGTCTTGCCAGAGTAC	2156
Oy	932	CAACACTGTCATGATGTATGATGCGGCCCTCTAGACCTACCCAGCGGACACATGCTT	991
Db	2157	CGACACGTGACTCTGACCAATACAGT-ACCCTCTTTACTCTGCCGCCACGGCGGACACAGCCT	2215
Oy	992	CCGGTA 997	
Db	2216	CCTGTA 2221	

```

RESULT 7
US-10-066-543-2864
, Sequence 2864, Application US/10066543
, Publication No. US20030087818A1
, GENERAL INFORMATION:
APPLICANT: Jiang, Yugu
APPLICANT: Pyle, Ruth A.
APPLICANT: Xu, Jiangchun
APPLICANT: Indirias, Carol yoseph
APPLICANT: Lodes, Michael J.
APPLICANT: Secrist, Heather
APPLICANT: Carter, Darrick
APPLICANT: Fanger, Gary R.
APPLICANT: Smith, Carole L.
APPLICANT: Durham, Margareta
APPLICANT: Stolk, John A.
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
TITLE OF INVENTION: AND DIAGNOSIS OF COLON CANCER
FILE REFERENCE: 210121.563
CURRENT APPLICATION NUMBER: US/10/066,543
CURRENT FILING DATE: 2002-01-31
NUMBER OF SEQ ID NOS: 3417
, SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 2864
LENGTH: 594
TYPE: DNA
ORGANISM: Homo sapiens
US-10-066-543-2864

```

Query Match	19.28;	Score 195.2;	DB 14;	Length 594;
Best Local Similarly	70.98;	Pred. No. 1.4e-52;		
Matches 373; Conservative	0;	Mismatches 113;	Indels 40;	Gaps 7;

```

OY 505 GAATAAAAGTCTCAGGTAAGCTTTTAAAGTTATTTCTGTCAAA--CATAGTG 562
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 6  GAAAGAAAGTCTCAGGGTAAACCTAAAGGTATTTCTGTCAATGCATAGTATG 65
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 563 ACAAAAT-----GGCGATCTTATCAGTGTCTT-----CTTGAGCCCCCTTGCCC 608
      |  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||
Db 66  ATGAAAGAAAGGTGGTATATCAGGAAGTTTCTTAAAGCTTTTCCTTCTCTACAC 125
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 609 CTGCTGTCCCTCCCAAGATGGGGGGTGTAGTCATATTAACTGGCCATCCTACAGTT 668
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 126  TCGCACAATGCCCAATTTGGGCAATTAAATTCATCTTAAACGTGTGTCTGTATGTC 185
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 669 GCTAACTTAGCAATGCTTTCTTTAGGACCCCTTCTTAAAGCAATATGTGTGACT 728
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

Db	186	GCTAACTGTAGTAGTCTTTTCTTCTTAAAGAACCCCTTCGACTGAGCAATAGCCT	-CCTT	244
Qy	729	GTCTCTTAAGACTTCTTCGATAATGACATTCGAGAGATTTTTTGGTAGATAGTAGAGAGTC	7888	
Db	245	GTATTATTAATAATCTTTCTGATATATGCATTAGAAAGTTTTTTTTTGGCATTTAGTAAATATGTC	3040	
Qy	789	GTTCTGTTTTCACCTTCCTTACACAGTCGACTAGTGCCTCCCTCGTTTTCAGTAATAC	848	
Db	305	TTTCAGATGTAC-----TTTATTACAGAGCTAATAAGTCTCTTCTTGTAGTTTACATAC	359	
Qy	849	TGGGTGTAAATATACAGTGTGCGGCTTTTACAGTTTTTAACTATTTTAAGTA-----	901	
Db	360	TAGGTGTAAATAATCATGTTGTGCACGCTTATATGTTTTTAAATATTTTAAATAATTTCTTA	419	
Qy	902	-----TTCTGAATCATCACTGTCTTGGCAGAGTACACACACTGTCTATGTATGA	951	
Db	420	AACATATGAACCTTCTTAAATCATCAGTCTGTTCGCATATTACGACACAGTGTCACTGTGACCA	479	
Qy	952	TGGCGGCCCTCTAGACGCTCACGACAGGGAGAACATGCTTCCGGTA	997	
Db	480	TACTG-AACCTCTTTACCTGTGGCCAGGGGAGAACGCGCTCTCTGTA	524	

```

RESULT 8
US-09-878-178-1654/c
: Sequence 1654, Application US/09878178
: Patent No. US20020177552a1
: GENERAL INFORMATION:
: APPLICANT: Jiang, Yuqiu
: APPLICANT: Harlocker, Susan L.
: APPLICANT: Secrist, Heather
: TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
: TITLE OF INVENTION: AND DIAGNOSIS OF COLON CANCER
: FILE REFERENCE: 210121.527
: CURRENT APPLICATION NUMBER: US/09/878,178
: CURRENT FILING DATE: 2001-06-08
: NUMBER OF SEQ. ID NOS: 2237
: SOFTWARE: FastSeq for Windows Version 4.0
: SEQ ID NO 1654
: LENGTH: 602
: TYPE: DNA
: ORGANISM: Homo sapien
: FEATURE:
: NAME/KEY: misc_feature
: LOCATION: (1)...(602)
: OTHER INFORMATION: n = A,T,C or G
US-09-878-178-1654

```

Query Match	19.0%	Score 193.2	DB 10	Length 602
Best Local Similarity	70.5%	Pred. No. 6.2e-52		
Matches 371	0	Mismatches 115	Indels 40	Gaps 7

OY	505	GAATAAAAGTCAAGGGTAAGCTTTTAAAGGTTAATTCCTGCAAA--CACTACATG	562
Db	592	GAATGAAGTCTCANGGGTAANAACCTAAAGGTAATTTCTGCAAAAGCATAGATG	533
OY	563	ACAAAT---GGCCGATCTTATCAGTGTCT-----CTTGAAGCCCCCTTCCC	608
Db	532	ATGAANAAGAGTGTGATTTATCAGAAATGTTTCTTAAAGCTTTTCCTTCTCTACAC	473
OY	609	CTGCTGCCCTCCCCAGATGGGGGTGAGTGCATATTAACTGGCATCTCTACAGTT	668
Db	472	CTGCATGCGCTCCCCAAATTTGGCATTAATTCATCTTAAACGAGTGTCTCTTATGTC	413
OY	669	GCTAACTTAGCAATGCTTTTCTTTAGACCCCTTTAAAGGCAATATGTGACT	728
Db	412	GCTAACTTAGAAATGCTTTCTATTAGAACCCCTTGACTGAGCAATATAGCT--CCTT	354
OY	729	GTAATAAAGATCTTTCTGATAAAGCAATGCGAATTTTTTGGTAGATAGTAAAGTGC	788
Db	353	GTATTATAAATCTTCTGATAAAGCAATTAAGAGTTTTTTTGCGATTAGTAAATATGC	294
OY	789	GTTCCTGTTTACACTTCCTTACTAGCTAGTACGTCCCTTGATTTCTAGTAAC	848







```
FILE REFERENCE: CCDNA-260XX
CURRENT APPLICATION NUMBER: US/09/871.161
CURRENT FILING DATE: 2001-05-31
PRIOR APPLICATION NUMBER: 09/328.111
PRIOR FILING DATE: 1999-06-08
PRIOR APPLICATION NUMBER: 60/117.393
PRIOR FILING DATE: 1999-01-27
PRIOR APPLICATION NUMBER: 60/098.639
PRIOR FILING DATE: 1998-08-31
NUMBER OF SEQ ID NOS: 544
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 483
LENGTH: 603
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc-feature
LOCATION: (1)...(603)
OTHER INFORMATION: n = A,T,C or G
US-09-871-161-483
```

```
Query Match 18.4%; Score 187.2; DB 11; Length 603;
Best Local Similarity 65.5%; Pred. No. 5.6e-50;
Matches 348; Conservative 0; Mismatches 158; Indels 25; Gaps 5;
```

```
OY 485 GTACATGAGTAATTCACCTGTAATTAAGTGTACGGGTAAAGCTTTTAACGGTTAAT 544
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 597 GGACAAGAAANNAAGTGGCGGTAAACCTTTAANGTTAATTTTGTCAATTCACATA 538
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
OY 545 TCTGTCAACAGTATG-ATGACAATGGCCGATCTTATGATGCTCTCTTGAAGCCCT 603
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 537 GATTAANNAANNAAGTTGATTAACCAAGNAATGTTTCTTANGCTTTTCTTNTCTT 478
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
OY 604 TCCCCCTGCTCCCTCCAGATGGGCGTTGATGTCATTAATTAACGGCCATCCCA 663
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 477 AACACCTGCTGCTGCTCCCAATGGGCGATTTAATTCATTTAAGNGTTTCTTCT 418
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
OY 664 CAGTGTACTTAAGCAAGTCTTTCTTTAGAACCCCTTCTTAACGCAATATGCT 723
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 417 TAGTCGCTAACTAGTAACTGCTTTCTTTATAGAACCCCTTCTGACGCAATATGCT 358
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
OY 724 GACCTGACTTAATGATCTTCTGATATGATTCGAGATTTTGTGATATGATA 783
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 357 -CCTGTATTAATTAATCTTTCTGATATGATTAAGAGTTTCTTGCATTAATA 299
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
OY 784 AGTGGCTTCCGTTTTCACCTTCTTACCTGAGTACGCTTCCCTTCCGTTTCTA 843
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 298 AGTCTTTCATGTTAC-----TTTATTCAGAGCTAATAAGTCTTCTTCTTCTA 244
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
OY 844 GTAACGGGTGTAGAAATCAGCTGCGGCTTTACAGTTTAACTAATTTAGATA-- 901
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 243 GTAACGTAGGTAAATATCATGTGTGAGCTTTATAGTTTAAATTTTATAGATA 184
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
OY 902 -----TCTGAACATCAGTCTTCCAGAGTACCAACATGTCATGTG 946
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 183 TCTTAACCTAAGAACTCTTAACTACCTCTTCCAGATTAACCGACATGTCATCTG 124
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
OY 947 ATTGATGCGCGCCCTCTAGACTCACCACGCGAGACATGCTCCGGTA 997
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 123 ACCAATGCTG-ACCCTCTTACCTGCGCCACGCGACACAGCCCTCTGTA 74
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
```

```
RESULT 14
US-10-066-543-3022/c
Sequence 3022, Application US/10066543
Publication No. US20030087818a1
GENERAL INFORMATION:
APPLICANT: Jjiang, Yugu
APPLICANT: Pyle, Ruth A.
APPLICANT: Xu, Jiangchun
APPLICANT: Indrias, Carol Joseph
APPLICANT: Lodes, Michael J.
APPLICANT: Secrist, Heather
```

```
APPLICANT: Carter, Darrick
APPLICANT: Fanger, Gary R.
APPLICANT: Smith, Carole L.
APPLICANT: Durham, Margarita
APPLICANT: Stolk, John A.
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
TITLE OF INVENTION: AND DIAGNOSIS OF COLON CANCER
FILE REFERENCE: 210121.563
CURRENT APPLICATION NUMBER: US/10/066.543
CURRENT FILING DATE: 2002-01-31
NUMBER OF SEQ ID NOS: 3417
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 3022
LENGTH: 477
TYPE: DNA
ORGANISM: Homo sapiens
US-10-066-543-3022
```

```
Query Match 18.2%; Score 185.4; DB 14; Length 477;
Best Local Similarity 73.2%; Pred. No. 1.8e-49;
Matches 301; Conservative 0; Mismatches 86; Indels 24; Gaps 4;
```

```
OY 604 TCCCCCTGCTCCCTCCAGATGGGCGTTGAGTCCATATTAAGTCCATCTCA 663
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 477 TACACCTGCTGCTCCCAATGGGCGATTTAATTCATCTTAACGTGTTCTCT 418
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
OY 664 CAGTGTACTTAAGCAAGTCTTTCTTTAGAACCCCTTCTTAACGCAATATGCT 723
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 417 TAGTCGCTAACTAGTAACTGCTTTCTTTATAGAACCCCTTCTGACGCAATATGCT 358
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
OY 724 GACCTGACTTAATGATCTTCTGATATGATTCGAGATTTTGTGATATGATA 783
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 357 -CCTGTATTAATTAATCTTTCTGATATGATTAAGAGTTTCTTCTCATTAATA 299
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
OY 784 AGTGGCTTCCGTTTTCACCTTCTTACGAGTACGATGCTTCCCTTCTTCTA 843
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 298 AGTCTTTCATGTTAC-----TTTATTCAGAGCTAATAAGTCTTCTTCTTCTA 244
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
OY 844 GTAACGGGTGTAGAAATCAGCTGCGGCTTTACAGTTTAACTAATTTAGATA-- 901
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 243 GTAACGTAGGTAAATATCATGTGTGAGCTTTTATAGTTTAAATTTTATAGATA 184
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
OY 902 -----TCTGAACATCAGTCTTCCAGAGTACCAACATGTCATGTG 946
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 183 TCTTAACCTAAGAACTCTTAACTACCTGCTTCCAGATTAACCGACATGTCATCTG 124
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
OY 947 ATTGATGCGCGCCCTCTAGACTCACCACGCGAGACATGCTCCGGTA 997
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 123 ACCAATGCTG-ACCCTCTTACCTGCGCCACGCGACACAGCCCTCTGTA 74
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
```

```
RESULT 15
US-09-960-352-7612
Sequence 7612, Application US/09960352
Patent No. US20020137139A1
GENERAL INFORMATION:
APPLICANT: Warren, Wesley C.
APPLICANT: Tao, Nengping
APPLICANT: Byatt, John C.
APPLICANT: Mathialagan, Nagappan
TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND
FILE REFERENCE: 16511.006/37-21(10298)C
CURRENT APPLICATION NUMBER: US/09/960.352
CURRENT FILING DATE: 2001-09-24
NUMBER OF SEQ ID NOS: 15112
SEQ ID NO 7612
LENGTH: 432
TYPE: DNA
ORGANISM: Bos taurus
OTHER INFORMATION: Clone ID: 33-LIB188-020-Q1-EL-A2
US-09-960-352-7612
```

Query Match	Similarity	16.9%	Score 172.2	DB 10	Length 432
Best Local	Similarity 72.2%		Pred. No. 3.4e-45		
Matches 294	Conservative 0		Mismatches 105	Indels 10	Gaps 5
QY	486	TACATGGAGTAATTCACACTGATATATAAAGTGTCACGGGTAAAGCTTTTAAAGGTTAAATTT	545		
DB	27	TACAGTAAATCTTAAACTATTTAAACATATCAGCGGTAAAGTTTAAAGTTAAATTT	86		
OY	546	CTGTCAAA - CAGTAGATGACAAATGGCGGATCTTATCAGTGTCTCTCTTGAGCCCC -	600		
DB	87	CTGTCAAAATGGGGTAGATGAAAGAAAGTCGGATTTATAGTAGAGTGTCTTAAAGCTTT	146		
OY	601	-CTTCCCTCTGTCTGTCCCTCCAGATGGGGCGTTGATGTCATATTTAACTGGCCATC	659		
DB	147	TCCTTAAACGTGCCACCCCTTGTGAATTTGGGCATTTATCATCTCTGACCTGGTCATC	206		
OY	660	CTCACAGTTGCTAACTTAGCAGAGTCTTTCTTTAGAACCCCTTTAACGAGCAATAT	719		
DB	207	CTCATAGTGGTAACTTAGTAGTCTTTCTTATAGAACATCATCTTAAAGGCAATAT	266		
OY	720	GTCGACCTGTACATAGATCTTTCGATATATGATTCGAGATTTTGTGTGATAG	779		
DB	267	GCCT - CCTGTATTTATTAATCTTTCTCAATATGATATAGAAATTTTGTAGTTAG	325		
OY	780	TAGAAGTCGTTCTGTTTACCTTCCCTTACT - AGCTAGACATAGCTTCCCTGGT	837		
DB	326	TAAAGTGCACCTCTTTTCATCTGTGGTTTATTCAAAAGCTAAATTAAGTCTTCTTGT	385		
OY	838	TTTCTAGTAATCT - GGGGTAGAAATCAAGTGTGGCGCTTACAGATT	883		
DB	386	TTTCTAGTAATCTTTAAATAAACAGTGTGGAGCTTTTATGGTT	432		

Search completed: August 26, 2003, 02:45:56  
Job time : 253.548 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 25, 2003, 21:56:10 ; Search time 1283.94 Seconds

(without alignments)  
11088.207 Million cell updates/sec

Title: US-09-717-321A-17

Perfect score: 348

Sequence: 1 tgaacatcactgtcttgc.....cttgaacttgaaaaaa 348

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 2888711 seqs, 2045481386 residues

Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Genbankl:\*

1: gb\_da:\*

2: gb\_hcg:\*

3: gb\_in:\*

4: gb\_om:\*

5: gb\_ov:\*

6: gb\_pat:\*

7: gb\_ph:\*

8: gb\_pl:\*

9: gb\_pr:\*

10: gb\_ro:\*

11: gb\_sts:\*

12: gb\_sy:\*

13: gb\_un:\*

14: gb\_vl:\*

15: em\_da:\*

16: em\_fun:\*

17: em\_hum:\*

18: em\_in:\*

19: em\_mu:\*

20: em\_om:\*

21: em\_or:\*

22: em\_ov:\*

23: em\_pat:\*

24: em\_ph:\*

25: em\_pl:\*

26: em\_ro:\*

27: em\_sts:\*

28: em\_un:\*

29: em\_vl:\*

30: em\_htg\_hum:\*

31: em\_htg\_inv:\*

32: em\_htg\_other:\*

33: em\_htg\_mus:\*

34: em\_htg\_pln:\*

35: em\_htg\_rnd:\*

36: em\_htg\_mam:\*

37: em\_htg\_vtl:\*

38: em\_sy:\*

39: em\_htgo\_hum:\*

40: em\_htgo\_mus:\*

41: em\_htgo\_other:\*

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query			DB ID	Description
		Match	Length	DB		
1	348	100.0	348	6	AX163753	AX163753 Sequence
2	348	100.0	1266	6	AX163738	AX163738 Sequence
3	320	92.0	217700	2	AC106124	AC106124 Rattus no
4	230.2	66.1	2281	10	BC051053	BC051053 Mus muscu
5	228.2	65.6	2319	10	BC003828	BC003828 Mus muscu
6	221	63.5	269081	2	AC068493	AC068493 Mus muscu
7	216.8	62.3	192498	2	AC105979	AC105979 Mus muscu
8	216.8	62.3	230015	2	AC132602	AC132602 Mus muscu
9	207.4	58.6	2302	9	BC050687	BC050687 Homo sapi
10	203.6	58.5	28567	9	HSN132695	HSN132695 Homo sapi
11	203.6	58.5	212827	9	AC009412	AC009412 Homo sapi
12	192.4	55.3	2315	9	AK054993	AK054993 Homo sapi
13	187.8	54.0	352	6	AX163754	AX163754 Sequence
14	187.4	53.9	1232	6	AX408985	AX408985 Sequence
15	187.4	53.9	1232	9	HDMPO2ST9	D25274 Homo sapien
16	176.4	50.7	87616	2	AC139405	AC139405 Homo sapi
17	176.4	50.7	137625	9	AC104663	AC104663 Homo sapi
18	175.6	50.5	240973	2	AC123247	AC123247 Rattus no
19	175.6	50.5	241048	2	AC129824	AC129824 Rattus no
20	174.8	50.2	5544	9	AF542527	AF542527 Homo sapi
21	170.6	49.0	262	11	G31709	G31709 swiss2233 Er
22	154	44.3	455	11	G26995	G26995 human STS S
23	123	35.3	123	6	AX163737	AX163737 Sequence
24	121.4	34.9	228121	2	AC133022	AC133022 Rattus no
25	121.4	34.9	239768	2	AC132582	AC132582 Rattus no
26	121.4	34.9	245668	2	AC130746	AC130746 Rattus no
27	116.4	33.4	110816	9	AC002404	AC002404 Human Chr
28	113	32.5	1017	6	AX163751	AX163751 Sequence
29	110.6	31.8	101584	9	CNS01DS5	AL121655 BAC seque
30	110.6	31.8	155943	2	AC012364	AC012364 Homo sapi
31	84.6	24.3	174316	2	AC022647	AC022647 Homo sapi
32	84.6	24.3	217249	9	AC009902	AC009902 Homo sapi
33	81.2	23.3	118648	2	AL138742	AL138742 Human DNA
34	81.2	23.3	131095	2	AL590071	AL590071 Homo sapi
35	79.2	22.8	170839	2	AC133467	AC133467 Mus muscu
36	75.2	21.6	237985	2	AC125754	AC125754 Rattus no
37	75.2	21.6	247478	2	AC097964	AC097964 Rattus no
38	72	20.7	156879	10	AL626786	AL626786 Mouse DNA
39	67	19.3	340701	2	AC120633	AC120633 Rattus no
40	66.8	19.2	62656	2	AC100110	AC100110 Mus muscu
41	66.8	19.2	178482	2	AC101810	AC101810 Mus muscu
42	60.8	17.5	64	6	AX522492	AX522492 Sequence
43	53.4	15.3	100697	5	AL672072	AL672072 zebrafish
44	52.4	15.1	162	6	AX397334	AX397334 Sequence
45	52.4	15.1	1022	6	AX163752	AX163752 Sequence

#### ALIGNMENTS

RESULT 1

AX163753

LOCUS AX163753 348 bp DNA linear PAT 22-JUN-2001

DEFINITION Sequence 17 from Patent WO0138579.

ACCESSION AX163753

VERSION AX163753.1 GI:14544859

KEYWORDS

SOURCE Rattus norvegicus (Norway rat)

ORGANISM Rattus norvegicus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;

Rattus.

REFERENCE 1

AUTHORS Gould-Rothberg, B.E., DiPippo, V.A., Ramsehl, T.M. and Gerweil, R.W.

TITLE Method of Identifying Toxic Agents using nsaid-induced differential

gene expression in liver  
 Patent: WO 0138579-A 17 31-MAY-2001;  
 Curagen Corporation (US)  
 Location/Qualifiers  
 FEATURES  
 source  
 1.348  
 /organism="Rattus norvegicus"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:10116"

BASE COUNT 103 a 76 c 67 g 102 t

ORIGIN

Query Match 100.0%; Score 348; DB 6; Length 348;  
 Best Local Similarity 100.0%; Pred. No. 1.8e-78;  
 Matches 348; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TGAACATCATCTCTTGGCAGAGTACCAACACTGTCTATGTGATGCCGCCCTCT 60  
 |||||||  
 Db 1 TGAACATCATCTCTTGGCAGAGTACCAACACTGTCTATGTGATGCCGCCCTCT 60  
 |||||||

QY 61 AGACCTACCCACGCGGACACATGCTCCGGTACCTTTGGGTCTGTGAGGTTCTGCAAG 120  
 |||||||  
 Db 61 AGACCTACCCACGCGGACACATGCTCCGGTACCTTTGGGTCTGTGAGGTTCTGCAAG 120  
 |||||||

QY 121 CGCTAGTGTCTAACGCCGTTCTGTACAACCTTACTCTGCGAAGAACACAGTGTGGGC 180  
 |||||||  
 Db 121 CGCTAGTGTCTAACGCCGTTCTGTACAACCTTACTCTGCGAAGAACACAGTGTGGGC 180  
 |||||||

QY 181 TTTCGACCACTAGAACAACTTTTCAATTGACAGTTGCGAATTTGGAGTGTGTTTA 240  
 |||||||  
 Db 181 TTTCGACCACTAGAACAACTTTTCAATTGACAGTTGCGAATTTGGAGTGTGTTTA 240  
 |||||||

QY 241 CATGATCTTTTCTTAATGAGTATGACAGTATTTGCTATGACTTAATAATCCCT 300  
 |||||||  
 Db 241 CATGATCTTTTCTTAATGAGTATGACAGTATTTGCTATGACTTAATAATCCCT 300  
 |||||||

QY 301 TGAATCATATAAAAAAAAAAATGTCTTTGGAATTTGAAAAAAA 348  
 |||||||  
 Db 301 TGAATCATATAAAAAAAAAAATGTCTTTGGAATTTGAAAAAAA 348  
 |||||||

RESULT 2  
 LOCUS AX163738 1266 bp DNA linear PAT 22-JUN-2001  
 DEFINITION Sequence 2 from Patent WO0138579.  
 AX163738  
 VERSION AX163738.1 GI:14544844  
 KEYWORDS  
 SOURCE Rattus norvegicus (Norway rat)  
 ORGANISM Rattus norvegicus  
 Eukaryota; Metazoa; Chordata; Cranialia; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;  
 Rattus.

REFERENCE  
 1 Gould-Rothberg, B.E., Dipippo, V.A., Ramsehl, T.M. and Gerwehl, R.W.  
 Method of identifying toxic agents using nsaid-induced differential  
 gene expression in liver  
 Patent: WO 0138579-A 2 31-MAY-2001;  
 Curagen Corporation (US)

FEATURES  
 source  
 1.1266  
 /organism="Rattus norvegicus"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:10116"

BASE COUNT 385 a 258 c 285 g 338 t

ORIGIN

Query Match 100.0%; Score 348; DB 6; Length 1266;  
 Best Local Similarity 100.0%; Pred. No. 2e-78;  
 Matches 348; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TGAACATCATCTCTTGGCAGAGTACCAACACTGTCTATGTGATGCCGCCCTCT 60  
 |||||||  
 Db 357 TGAACATCATCTCTTGGCAGAGTACCAACACTGTCTATGTGATGCCGCCCTCT 298  
 |||||||

QY 61 AGACCTACCCACGCGGACACATGCTCCGGTACCTTTGGGTCTGTGAGGTTCTGCAAG 120  
 |||||||  
 Db 297 AGACCTACCCACGCGGACACATGCTCCGGTACCTTTGGGTCTGTGAGGTTCTGCAAG 238  
 |||||||

QY 121 CGCTAGTGTCTAACGCCGTTCTGTACAACCTTACTCTGCGAAGAACACAGTGTGGGC 180  
 |||||||  
 Db 237 CGCTAGTGTCTAACGCCGTTCTGTACAACCTTACTCTGCGAAGAACACAGTGTGGGC 178  
 |||||||

QY 181 TTTCGACCACTAGAACAACTTTTCAATTGACAGTTGCGAATTTGGAGTGTGTTTA 240  
 |||||||  
 Db 177 TTTCGACCACTAGAACAACTTTTCAATTGACAGTTGCGAATTTGGAGTGTGTTTA 118  
 |||||||

QY 241 CATGATCTTTTCTTAATGAGTATGACAGTATTTGCTATGACTTAATAATCCCT 300  
 |||||||  
 Db 117 CATGATCTTTTCTTAATGAGTATGACAGTATTTGCTATGACTTAATAATCCCT 58  
 |||||||

QY 301 TGAATCATATAAAAAAAAAAATGTCTTTGGAATTTGAAAAAAA 348  
 |||||||  
 Db 57 TGAATCATATAAAAAAAAAAATGTCTTTGGAATTTGAAAAAAA 10  
 |||||||

RESULT 3  
 LOCUS AC106124 217700 bp DNA linear HTG 10-MAY-2003  
 DEFINITION Rattus norvegicus clone CH230-119E11, \*\*\* SEQUENCING IN PROGRESS  
 AC106124  
 VERSION AC106124.6 GI:30521557  
 KEYWORDS HTG: HTGS\_PHASE1; HTGS\_DRAFT; HTGS\_ENRICHED.  
 SOURCE Rattus norvegicus (Norway rat)  
 ORGANISM Rattus norvegicus  
 Eukaryota; Metazoa; Chordata; Cranialia; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;  
 Rattus.

1 (bases 1 to 217700)  
 Muzny, D., Marle, Metzker, M., Lee, Abramson, S., Adams, C., Alder, J.,  
 Allen, C., Allen, H., Alsbrooks, S., Amin, A., Anguiano, D.,  
 Anyalebech, V., Aoyagi, A., Ayodeji, M., Baca, E., Baden, H.,  
 Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benham, F.,  
 Bismah, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M.,  
 Bryant, N., Buhay, C., Burch, P., Burrell, K., Calderon, E.,  
 Cardenas, V., Carter, K., Cavazos, I., Caesar, H., Center, A.,  
 Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Y., Chen, Z., Chu, J.,  
 Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Grege, A., D'Souza, L.,  
 Davila, M., Davis, C., Davy-Carroll, L., De Anda, C., Dedrich, D.,  
 Delgado, O., Denson, S., Deramo, C., Ding, Y., Dinh, H., Ditya, K.,  
 Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Eaves, K.,  
 Egan, A., Escotto, M., Eugene, C., Evans, C.A., Falls, T., Fan, G.,  
 Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P.,  
 Fraser, C.M., Gabisi, A., Ganta, R., Garcia, A., Garner, T., Garza, M.,  
 Gebregeorgis, E., Geer, K., Gill, R., Girdy, M., Guerra, W., Guevara, W.,  
 Gunaratne, P., Haaland, W., Hamill, C., Hamilton, C., Hamilton, K.,  
 Harvey, Y., Havlak, P., Hawes, A., Henderson, N., Hernandez, J.,  
 Hernandez, R., Hines, S., Hladun, S.L., Hodgson, A., Hogue, M.,  
 Hollins, B., Howells, S., Hulyk, S., Hume, J., Idelbird, D., Jackson, A.,  
 Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolyet, A.,  
 Karpathy, S., Kelly, S., Kelly, S., Khan, S., King, R., Kovar, C.,  
 Kowis, C., Kraft, C.L., Ledow, H., Levan, J., Lewis, L., Li, Z., Liu, J.,  
 Liu, J., Liu, M., Liu, Y., London, P., Longacre, S., Lopez, J.,  
 Lorensuewa, L., Loulsegod, H., Lozano, R.J., Lu, X., Ma, J.,  
 Maheshwari, M., Mahlender, M., Mahmoud, M., Malloy, K., Mangum, A.,  
 Mangum, B., Mapua, P., Martin, K., Martin, R., Martinez, E.,  
 Mawhney, S., McLeod, M.P., McNeill, T.Z., Meenan, E.,  
 Milosavljevic, A., Miner, G., Minja, E., Montemayor, J., Moore, S.,  
 Morgan, M., Morris, K., Morris, S., Munidasa, M., Murphy, M., Nair, L.,  
 Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S.,  
 Nwankweme, O., Okwuonu, G., Olarinmoye, A., Pal, S., Parks, K.,  
 Pasternak, S., Paul, H., Perez, A., Perez, L., Pfannkuch, C.,  
 Plopper, F., Poindexter, A., Popovic, D., Primus, E., Pu, L.,  
 Puzo, M., Quiroz, J., Rachlin, E., Reeves, K., Regier, R., Reigh, R.,  
 Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F.,  
 Rives, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S.J.,

Sanders, W., Saverly, G., Scherer, S., Scott, G., Shatsman, S., Shen, H., Shetty, J., Shvartsbeyn, A., Sisson, I., Sitter, C.D., Smajs, D., Sneed, A., Sodergren, E., Song, X.-Z., Sorelle, R., Sosa, J., Steimle, M., Strong, R., Sutton, A., Svatek, A., Taber, P., Taylor, C., Taylor, T., Thomas, R., Thomas, S., Tingey, A., Trejos, Z., Umatt, K., Valas, R., Vera, V., Villasana, D., Waldron, L., Walker, B., Wang, J., Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, F., Williams, G., Willson, R., Wleczek, R., Wooden, H., Worley, K., Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V., Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von Niederhausern, A., Weiss, R., Smith, D.R., Holt, R.A., Smith, H.O., Weinstock, G., and Gibbs, R.A.

Unpublished  
2 (bases 1 to 217700)  
Worley, K.C.

Direct Submission  
Submitted (12-JAN-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA  
3 (bases 1 to 217700)  
Rat Genome Sequencing Consortium.

Direct Submission  
Submitted (10-MAY-2003) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA  
On May 10, 2003 this sequence version replaced g1:25094619. The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (http://www.hgsc.bcm.tmc.edu/projects/atlas/). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.

----- Genome Center  
Center: Baylor College of Medicine  
Center code: BCM  
Web site: http://www.hgsc.bcm.tmc.edu/  
Contact: hgsc-help@bcm.tmc.edu  
----- Project Information  
Center project name: CH230-119E11  
Center clone name: CH230-119E11  
----- Summary Statistics  
Assembly program: Atlas 3.0;  
Consensus quality: 185725 bases at least Q40  
Consensus quality: 190140 bases at least Q30  
Consensus quality: 192842 bases at least Q20  
Estimated insert size: 194758; sum-of-contigs estimation  
Quality coverage: 5x in Q20 bases; sum-of-contigs estimation  
-----  
\* NOTE: Estimated insert size may differ from sequence length  
\* (see http://www.hgsc.bcm.tmc.edu/docs/genbank\_drift\_data.html).  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 11 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

1 109797: contig of 109797 bp in length  
\* 109798 109897: gap of unknown length  
\* 109898 166934: contig of 57037 bp in length  
\* 166935 167034: gap of unknown length  
\* 167035 176384: contig of 9350 bp in length  
\* 176385 176484: gap of unknown length  
\* 176485 206327: contig of 29843 bp in length  
\* 206328 206427: gap of unknown length

206428 207461: contig of 1034 bp in length  
\* 207462 207561: gap of unknown length  
\* 207562 208607: contig of 1046 bp in length  
\* 208608 208707: gap of unknown length  
\* 208708 209708: contig of 1201 bp in length  
\* 209709 210008: gap of unknown length  
\* 210009 211583: contig of 1575 bp in length  
\* 211584 211683: gap of unknown length  
\* 211684 213649: contig of 1966 bp in length  
\* 213650 213750: gap of unknown length  
\* 213750 215115: contig of 1366 bp in length  
\* 215116 215215: gap of unknown length  
\* 215216 217700: contig of 2485 bp in length.  
\* 215216  
Location/Qualifiers  
1. 217700  
source  
/organism="Rattus norvegicus"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:10116"  
/clone="CH230-119E11"

BASE COUNT 51384 a 45640 c 45954 g 51476 t 23246 others  
ORIGIN

Query Match 92.0%; Score 320; DB 2; Length 217700;  
Best Local Similarity 95.6%; Pred. No. 3.6e-71;  
Matches 329; Conservative 0; Mismatches 15; Indels 0; Gaps 0;

QY 1 TGAACATCAGCTGCTTGGCAGATGACCAACACTGTCATGTGATGATGCCGCCCTCT 60  
|||||  
DB 64351 TGAACATCAGCTGCTTGGCAGATGACCAACACTGTCATGTGATGATGCCGCCCTCT 64410  
|||||  
QY 61 AGACCTCACCCAGCGGAGACATGCTCCGGTACCTTGGGTGCTGTGAGTTCTGTCAAG 120  
|||||  
DB 64411 AGACCTCACCCAGCGGAGACATGCTCCGGTACCTTGGGTGCTGTGAGTTCTGTCAAG 64470  
|||||  
QY 121 CGCTAGTGTCTAACGCCCTTGTGTACACCTAATCTACATGCGCAAGACAGCTTGGGCC 180  
|||||  
DB 64471 CGCTAGTGTCTAACGCCCTTGTGTACACCTAATCTACATGCGCAAGACAGCTTGGGCC 64530  
|||||  
QY 181 TTTGACCACTGAGCAAACTTTTTCATTTGACAGTTGCAAAATGGAGTGTCTTTTA 240  
|||||  
DB 64531 TTTGACCACTGAGCAAACTTTTTCATTTGACAGTTGCAAAATGGAGTGTCTTTTA 64590  
|||||  
QY 241 CATGATCTTTTGTCTAATGACATGATGATTTTGCATGTATGACTTAATAATCT 300  
|||||  
DB 64591 CATGATCTTTTGTCTAATGACATGATGATTTTGCATGTATGACTTAATAATCT 64650  
|||||  
QY 301 TGAATCATTAATAAAAAAAAAAATGCTTTGGACCTTGAANA 344  
|||||  
DB 64651 TGAATCATTAATAAACTGTGATGCTTGTGTTGGAACTTGACAA 64694  
|||||

RESULT 4  
BC051053 2281 bp mRNA linear ROD 14-APR-2003  
LOCUS Mus musculus, clone MGC:58966 IMAGE:5038182, mRNA, complete cds.  
DEFINITION BC051053  
ACCESSION BC051053.1 GI:29835221  
VERSION MGC:  
KEYWORDS Mus musculus (house mouse)  
SOURCE Mus musculus  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
Strausberg, R.  
REFERENCE  
AUTHORS Direct Submission  
TITLE Submitted (11-APR-2003) National Institutes of Health, Mammalian  
Gene Collection (MGC), Cancer Genomics Office, National Cancer  
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,  
USA  
REMARK NIH-MGC Project URL: http://mgc.nci.nih.gov  
COMMENT Contact: MGC help desk  
Email: cgapbs-remail.nih.gov  
Tissue Procurement: Gilbert Smith, Ph.D.





```

gene
    /lab_host="DH10B"
    /note="Vector: PCMV-SPORT6"
    .2319
    /gene="Rac1"
    /note="synonym: D5ertd559e"
    /db_xref="LOCUSID:19353"
    /db_xref="MGI:97845"
    198..776
    /codon_start=1
    /product="Rac1 protein"
    /protein_id="AAH03828.1"
    /db_xref="GI:13277918"
    /db_xref="LocusID:19353"
    /translation="MQAIKCVAGDGVAVGKTKLLISYTNAPGPEYIPTVDNVSANV
    MYDKPVLGLMDPAGQEDYDLRPLSPQTDVFLISYTNAPGPEYIPTVDNVSANV
    HHCNPPIIVGLTDLDDKTEKLEKLTPTTYPOGLAMAEIGAVVLECSAL
    TORGLKVFDEAIRAVLCPPVKKRKRKCLL"

BASE COUNT      603 a      573 c      503 g      640 t

ORIGIN
Query Match      65.6%; Score 228.2; DB 10; Length 2319;
Best Local Similarity 83.7%; Pred. No. 8.5e-48;
Matches 292; Conservative 0; Mismatches 33; Indels 24; Gaps 2;

OY 21 AGAGTACCAACTGCTCAATGATGATGCCGCCCTCTAGACCTCACCCAGCGGACA 80
    |||||||
DB 1964 ACAGTACCAACTGCTCAATGATGATGCCGCCCTCTAGACCTCACCCAGCGGACA 2020
    |||||||

OY 81 CATGCTCCGGTA-----CCTTGGTCTGTGAGTCTCTGTA 119
    |||||||
DB 2021 GAGCTTCCTGTCAATGCTGCTGCTGCTGAGATGTTCTTGGGCTGTGAGTCTGTCAA 2080
    |||||||

OY 120 GCGCTAGTGTCAAGCCCTTCTGTACAACTACTACTGCGCAGACAGACAGTGTGGAC 179
    |||||||
DB 2081 CTGCTAGTGTCAAGCCCTTCTGTACAACTACTACTGCGCAGACAGTGTGGAC 2140
    |||||||

OY 180 CTTTCGACCACTAGCAAACTTTTTCATTTGACAGTTCGACAAATTTGGAGTGTGTT 239
    |||||||
DB 2141 CTTTCACCACTAGCAAACTTTTTCATTTGACAGTTCGACAAATTTGGAGTGTGTT 2200
    |||||||

OY 240 ACATTGATCTTTGCTAATGAGTATGATGTTTTCATGATGATGATTAATTAATCC 299
    |||||||
DB 2201 ACATTGATCTTTGCTAATGAGTATGATGATGTTTTCATGATGATTAATTAATCC 2260
    |||||||

OY 300 TTGATCATTAATAAAAAAAAAAATGCTTTGGACTGAAAAA 348
    |||||||
DB 2261 TTGATCATTAATAAAAAAAAAAATGCTTTGGACTGAAAAA 2309
    |||||||

RESULT 6
AC068493      269081 bp   DNA      linear   HTG 15-MAY-2002
LOCUS
DEFINITION
Mus musculus clone RP23-76k1 strain C57BL6/J, WORKING DRAFT
SEQUENCE, 26 unordered pieces.
ACCESSION
AC068493.10 GI:15148081
VERSION
HTG: HTGS_PHASE1; HTGS_DRAFT.
KEYWORDS
Mus musculus (house mouse)
SOURCE
ORGANISM
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE
AUTHORS
Montgomery,K.T., Grills,G., Han,J., Lee,E., Long,J., Pomerantz,R.,
Ioshikhes,I.P., Shih,C., Decker,J., Thomas,E., Peters,A.,
Gordon,M., Goltz,J.S. and Kucherlapati,R.
TITLE
High Throughput Mouse Sequencing
JOURNAL
Unpublished
2 (bases 1 to 269081)
REFERENCE
AUTHORS
Montgomery,K.T., Grills,G., Han,J., Lee,E., Long,J., Pomerantz,R.,
Ioshikhes,I.P., Shih,C., Decker,J., Thomas,E., Peters,A.,
Gordon,M., Goltz,J.S. and Kucherlapati,R.
TITLE
Direct Submission
JOURNAL
Submitted (03-MAY-2000) Department of Molecular Genetics, Albert

```

## COMMENT

```

Einstein College of Medicine Genome Center, 1300 Morris Park Ave.,
Bronx, NY 10461, USA
On Aug 11, 2001 this sequence version replaced gi:14993654.
-----Genome Center
Center: Harvard Partners Genome Center
Center Code: HPGC
Web site: http://www.hpcg.org/Sequence/mouse.html
Contact: hpgc@emdel.mgh.harvard.edu
-----Summary Statistics
Center project name: ABN
Sequencing vector: pUC18; L08752
Chemistry: Dye-terminator Big Dye; 100%
*Consensus quality: 256302 at least Q20
*Consensus quality: 254124 at least Q30
*Consensus quality: 250276 at least Q40
*Estimated insert size: agarose-FP - N/A
Quality coverage: agarose-FP - N/A
Quality coverage: 6.2 x in Q20 bases; sum-of-coverage estimation
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 26 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
1 51730: contig of 51730 bp in length
* 51731 51750: gap of unknown length
* 51751 82423: contig of 30673 bp in length
* 82424 82443: gap of unknown length
* 82444 118158: contig of 35715 bp in length
* 118159 118178: gap of unknown length
* 118179 139441: contig of 21263 bp in length
* 139442 139461: gap of unknown length
* 139462 160638: contig of 21177 bp in length
* 160639 176618: gap of unknown length
* 176619 176632: contig of 16953 bp in length
* 176633 190991: gap of unknown length
* 190992 191011: contig of 13360 bp in length
* 191012 200128: gap of unknown length
* 200129 200148: contig of 9117 bp in length
* 200149 211882: gap of unknown length
* 211883 211902: contig of 11734 bp in length
* 211903 220773: gap of unknown length
* 220774 220793: contig of 8871 bp in length
* 220794 229643: gap of unknown length
* 229644 229663: contig of 8850 bp in length
* 229664 238339: gap of unknown length
* 238340 238359: contig of 8676 bp in length
* 238360 244848: gap of unknown length
* 244849 244868: contig of 6489 bp in length
* 244869 249725: gap of unknown length
* 249726 249745: contig of 4857 bp in length
* 249746 252523: gap of unknown length
* 252524 252543: contig of 2778 bp in length
* 252544 257520: gap of unknown length
* 257521 257540: contig of 4977 bp in length
* 257541 260396: gap of unknown length
* 260397 260416: contig of 2856 bp in length
* 260417 262729: gap of unknown length
* 262730 262739: contig of 2293 bp in length
* 262740 263934: gap of unknown length
* 263935 263954: contig of 1205 bp in length
* 263955 263973: gap of unknown length
* 263974 264196: contig of 242 bp in length
* 264197 264217: gap of unknown length
* 264218 264629: contig of 413 bp in length
* 264630 264649: gap of unknown length
* 264650 265608: contig of 939 bp in length
* 265609 265628: gap of unknown length
* 265629 266935: contig of 1307 bp in length

```



Cook, A., Cooke, P., Dearellano, K., Dewar, K., Diaz, J.S., Dodge, S.,  
 Fato, S., Ferreira, P., Fitzgerald, M., Gage, D., Galagan, J.,  
 Gady, S., Gord, S., Graham, L., Grand-Pierre, N., Hafez, N.,  
 Hagos, B., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C.,  
 Kanat, A., Karatas, A., Kells, C., Landers, T., Levine, R.,  
 Lindblad-Toh, K., Liu, G., Maclean, C., MacDonald, P., Major, J.,  
 Matthews, C., McCarthy, M., Meldrum, J., Menus, L., Mihova, T.,  
 Mingna, V., Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Norbu, C.,  
 Norman, C.H., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J.,  
 Peterson, K., Phunkhang, P., Pierre, N., Raymond, C., Retta, R.,  
 Rise, C., Rogov, P., Roman, J., Roy, A., Schauer, S., Schupbach, R.,  
 Seaman, S., Severy, P., Smith, C., Spencer, B., Stange-Thomann, N.,  
 Stojanovic, N., Talamas, J., Testaye, S., Theodore, J., Topham, K.,  
 Travers, M., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X.,  
 Wyman, D., Young, G., Zainoun, J., Zemke, L., Zimmerman, A. and Zody, M.  
 Direct Submission  
 Submitted (21-OCT-2002) Whitehead Institute/MIT Center for Genome  
 Research, 320 Charles Street, Cambridge, MA 02141, USA  
 On Oct 21, 2002 this sequence version replaced gi:20149424.  
 All repeats were identified using RepeatMasker:  
 Smit, A.F.A. & Green, P. (1996-1997)  
 http://ftp.genome.washington.edu/RM/RepeatMasker.html

----- Genome Center  
 Center: Whitehead Institute/ MIT Center for Genome Research  
 Center code: WIBR  
 Web site: http://www.seq.wi.mit.edu  
 Contact: sequence\_submissions@genome.wi.mit.edu

----- Project Information  
 Center project name: L20199

Center clone name: 301\_E\_12

----- Summary Statistics  
 Sequencing vector: Plasmid; n/a; 100% of reads  
 Chemistry: Dye-terminator Big Dye; 100% of reads  
 Assembly program: Phrap; version 0.960731  
 Consensus quality: 190070 bases at least Q40  
 Consensus quality: 191243 bases at least Q30  
 Consensus quality: 191514 bases at least Q20  
 Insert size: 16300; agarose-fp  
 Insert size: 191698; sum-of-ctrls  
 Quality coverage: 11.6 in Q20 bases; agarose-fp  
 Quality coverage: 9.9 in Q20 bases; sum-of-ctrls

\* NOTE: This is a 'working draft' sequence. It currently  
 \* consists of 9 contigs. Gaps between the contigs  
 \* are represented as runs of N. The order of the pieces  
 \* is believed to be correct as given, however the sizes  
 \* of the gaps between them are based on estimates that have  
 \* provided by the submitter.

\* This sequence will be replaced  
 \* by the finished sequence as soon as it is available and  
 \* the accession number will be preserved.

\* 1 1061: gap of 1061 bp in length  
 \* 1062 1161: gap of 100 bp  
 \* 1162 5714: contig of 4553 bp in length  
 \* 5715 5814: gap of 100 bp  
 \* 5815 11422: contig of 5608 bp in length  
 \* 11423 11522: gap of 100 bp  
 \* 11523 17600: contig of 6078 bp in length  
 \* 17601 17700: gap of 100 bp  
 \* 17701 30157: contig of 12457 bp in length  
 \* 30158 30257: gap of 100 bp  
 \* 30258 38787: contig of 8530 bp in length  
 \* 38788 38887: gap of 100 bp  
 \* 38888 99444: contig of 60557 bp in length  
 \* 99445 99544: gap of 100 bp  
 \* 99545 143263: contig of 43719 bp in length  
 \* 143264 143364: gap of 100 bp  
 \* 143364 192498: contig of 49135 bp in length.

## FEATURES

source

1. 192498  
 /organism="Mus musculus"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:10090"

/clone="RP24-301E12"  
 /clone\_11b="RPC1-24 Male Mouse BAC"  
 1. 1061  
 /note="assembly\_fragment"  
 1162. 5714  
 /note="assembly\_fragment"  
 5815. 11422  
 /note="assembly\_fragment"  
 11523. 17600  
 /note="assembly\_fragment"  
 17701. 30157  
 /note="assembly\_fragment"  
 30258. 38787  
 /note="assembly\_fragment"  
 38888. 99444  
 /note="assembly\_fragment"  
 99545. 143263  
 /note="assembly\_fragment"  
 143364. 192498  
 /note="assembly\_fragment  
 clone\_end:17  
 vector\_slide:right"

BASE COUNT 53596 a 42471 c 43252 g 52379 t 800 others  
 ORIGIN

Query Match 62.3%; Score 216.8; DB 2; Length 192498;  
 Best Local Similarity 86.1%; Pred. No. 8.6e-45;  
 Matches 272; Conservative 0; Mismatches 22; Indels 22; Gaps 2;

QY 29 AACACTGTCATGTGATGTATGCGCCCTCTAGACCTCACCACGCGACATGCTTC 88  
 DB 90554 AACACTGTCATGTGATGTATGCGCTG-CCTCTAGACCTCACCACGACAGCGCTTC 90496

QY 89 CGGTA-----CCTTGGGCTGTGAGGTTCTCTCAAGCGCTGAT 127  
 DB 90495 CTGTCGTGGCTCTGCTAGAGATGTTCCCTGGGCTGTGAGGTTCTCTCAAGCGCTGAT 90436

QY 128 GCTAAGCGGCTTCTTACACCTACCTACCTACGCAAGAGTGGGCTTTCGAC 187  
 DB 90435 GCTAATGCTGTCCTGTACACCTACCTACCTACGCAAGATACAGTGGGCTTTCGAC 90376

QY 188 CACTGAGAACAACTTTTCAATTGACAGTTGCAGAAATGTGTGAGTGTTTTACATTTGAT 247  
 DB 90375 CACTGAGAACAACTTTTCAATTGACAGTTGCAGAAATGTGTGAGTGTTTTACATTTGAT 90316

QY 248 CTTTGTCAATGACCTTGACATGTTTGCATGCTATGCTTAATTAATCCTGAATCA 307  
 DB 90315 CTTTGTCAATGACCTTGACATGTTTGCATGCTATGCTTAATTAATCCTGAATCA 90256

QY 308 TAAAAAAGAAAAA 323  
 DB 90255 TTAAGAAAAAAGAAA 90240

## RESULT 8

AC132602 230015 bp DNA linear HTG 01-SEP-2002  
 AC132602/c Mus musculus chromosome UNK clone Rp23-477013, WORKING DRAFT

DEFINITION SEQUENCE, 10 unordered pieces.

AC132602.1 GI:22597688  
 VERSION HTG; HTGS\_PHASE1; HTGS\_DRAFT; HTGS\_FULLTOP.  
 KEYWORDS Mus musculus (house mouse)

## ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

## REFERENCE

1 (bases 1 to 230015)

## AUTHORS

McPherson, J.D. and Waterston, R.H.

## TITLE

The sequence of Mus musculus clone

## JOURNAL

Unpublished

## REFERENCE

2 (bases 1 to 230015)

## AUTHORS

McPherson, J.D. and Waterston, R.H.

## TITLE

Direct Submission

JOURNAL Submitted (01-SEP-2002) Genome Sequencing Center, 4444 Forest Park  
Parkway, St. Louis, MO 63108, USA

## COMMENT

----- Genome Center -----  
Center: Washington University Genome Sequencing Center  
Center code: WUGSC  
Web site: <http://genome.wustl.edu/gsc/index.shtml>  
Contact: [submissions@wustl.edu](mailto:submissions@wustl.edu)  
----- Project Information -----  
Center project name: M\_BA0477C13  
----- Summary Statistics -----  
Sequencing vector: MJ3: 0%  
Chemistry: Dye-Primer ET; 0% of reads  
Chemistry: Dye-terminator Big Dye; 100% of reads  
Assembly program: Phrap; version 0.990319  
Consensus quality: 225083 bases at least Q40  
Consensus quality: 225730 bases at least Q30  
Consensus quality: 226267 bases at least Q20  
Insert size: 20200; agarose-1p  
Insert size: 228405; sum-of-ctrls  
Quality coverage: 13.11 in Q20 bases; agarose-1p  
Quality coverage: 9.64 in Q20 bases; sum-of-ctrls

\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 10 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

```

1      710: contig of 710 bp in length
*      711      810: gap of unknown length
*      811      18316: contig of 17506 bp in length
*      18317      18416: gap of unknown length
*      18417      57380: contig of 38964 bp in length
*      57381      57480: gap of unknown length
*      57481      118173: contig of 60693 bp in length
*      118174      118273: gap of unknown length
*      118274      201518: contig of 83245 bp in length
*      201519      201618: gap of unknown length
*      201619      202801: contig of 1183 bp in length
*      202802      202901: gap of unknown length
*      202902      204714: contig of 1813 bp in length
*      204715      204814: gap of unknown length
*      204815      208091: contig of 3277 bp in length
*      208092      208191: gap of unknown length
*      208192      213082: contig of 4891 bp in length
*      213083      213182: gap of unknown length
*      213183      230015: contig of 16833 bp in length.

```

## FEATURES

## source

```

1. .230015
  /organism="Mus musculus"
  /mol_type="genomic DNA"
  /db_xref="taxon:10090"
  /chromosome="UNK"
  /clone="RP23-477C13"
  1. .710
    /note="assembly_name:Contig1"
    811. .18316
      /note="assembly_name:Contig10"
      18417. .57380
        /note="assembly_name:Contig11"
        57481. .118173
          /note="assembly_name:Contig12"
          118274. .201518
            /note="assembly_name:Contig13"
            201619. .202801
              /note="assembly_name:Contig3"
              202902. .204714
                /note="assembly_name:Contig6"

```

```

misc_feature      204815..208091      /note="assembly_name:Contig7"
misc_feature      208192..213082      /note="assembly_name:Contig8"
misc_feature      213183..230015      /note="assembly_name:Contig9"
BASE COUNT      63025 a 50551 c 50278 g 65213 t      948 others
ORIGIN

```

Query Match 62.3%; Score 216.8; DB 2; Length 230015;  
Best Local Similarity 86.1%; Pred. No. 8, 6e-45;  
Matches 272; Conservative 0; Mismatches 22; Indels 22; Gaps 2;

```

QY      29  AACACTGTCATGATGATGATGCCGCCCTTAGACCTCACCCGCGGACACATGCTTC 88
      |||||
DB      109281  AACACTGTCATGATGATGATGCTG-CCTCTGACCTCAGTCACGAGACAGACGCTTC 109223
      |||||
QY      89  CGGTA-----CCTTGGCTGAGGTGCTGCAAGCGGTACT 127
      |||||
DB      109222  CTCTCGTGGCTGCTGCTGACGATGCTTGGGCTGAGGTCTGCACTGCTACT 109163
      |||||
QY      128  GCTAACGCGTCTCTGACACCTACTACTGCAAGACACAGTGTGGCCTTCGAC 187
      |||||
DB      109162  GCTAATGCTGCTCTGTAACAACCTAATCACTGCGAAGATACGCTTGGCCTTCAC 109103
      |||||
QY      188  CACTAAGACAACTTTTTCATTTGACAGTTGACAGATTGTGAGTGTTCATATTGAT 247
      |||||
DB      109102  CACTAAGACAACTTTTTCAGTTGACAGTTGACAGATTGTGAGTGTTCATATTGAT 109043
      |||||
QY      248  CTTTTCCTAATGAGTATGATGATGATGATGATGATGATGATGATGATGATGATGAT 307
      |||||
DB      109042  CTTTTCCTAATGAGTATGATGATGATGATGATGATGATGATGATGATGATGATGAT 108983
      |||||
QY      308  TAAATAAAAAAAAAA 323
      |||||
DB      108982  TTAATAAAAAAAAAA 108967
      |||||

```

RESULT 9  
BC050687  
LOCUS  
DEFINITION  
BC050687 2302 bp mRNA linear PRI 11-APR-2003  
Homo sapiens, ras-related C3 botulinum toxin substrate 1 (rho  
family, small GTP binding protein Rac1), clone MGC:60264  
IMAGE:6149377, mRNA, complete cds.

ACCESSION  
BC050687.1 GI:29792301  
VERSION  
BC050687.1  
KEYWORDS  
MGC.  
SOURCE  
Homo sapiens (human)  
ORGANISM  
Homo sapiens

REFERENCE  
1 (bases 1 to 2302)  
Mammalia: Eutheria; Primates: Catarrhini; Homiidae; Homo.  
Strausberg, R.

AUTHORS  
TITLE  
JOURNAL  
SUBMITTED (08-APR-2003) National Institutes of Health, Mammalian  
Gene Collection (MGC), Cancer Genomics Office, National Cancer  
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,  
USA

REMARK  
NIR-MGC Project URL: <http://mgc.ncl.nih.gov>  
Contact: MGC help desk  
Email: [cgabs-remail.nih.gov](mailto:cgabs-remail.nih.gov)  
Tissue Procurement: ATCC/DCPD/DMP

CDNA Library Preparation: Life Technologies, Inc.  
DNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA sequencing by: Sequencing Group at the Stanford Human Genome  
Center, Stanford University School of Medicine, Stanford, CA 94305  
Web site: <http://www-shgc.stanford.edu>  
Contact: (Dickson, Mark) [mdépakxi@stanford.edu](mailto:mdépakxi@stanford.edu)  
Dickson, M., Schmutz, J., Grimwood, J., Rodriguez, A., and Myers,  
R. M.

Clone distribution: MGC clone distribution information can be found  
through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>



```
intron      1262..13700
            /gene="rac1"
            /number=1
repeat_region /rpt_family="GC_rich"
            1390..1419
repeat_region /rpt_family="FLAM"
            1752..1793
repeat_region /rpt_family="AluJo"
            1794..2005
repeat_region /rpt_family="AluJo"
            2140..2171
repeat_region /rpt_family="Simple-repeat"
            2666..2972
repeat_region /rpt_family="AluSx"
            2980..3267
repeat_region /rpt_family="AluSx"
            3268..3289
repeat_region /rpt_family="(r)n"
            3802..4102
repeat_region /rpt_family="AluJb"
            4122..4386
repeat_region /rpt_family="AluSx"
            4538..4719
repeat_region /rpt_family="LTR3A"
            4754..5053
repeat_region /rpt_family="AluSg"
            5297..5594
repeat_region /rpt_family="AluSx"
            5655..5757
repeat_region /rpt_family="AluSg"
            5758..6062
repeat_region /rpt_family="AluSg"
            6063..6198
repeat_region /rpt_family="AluSg"
            6222..6401
repeat_region /rpt_family="AluSg/x"
            7162..7238
repeat_region /rpt_family="LIMC5"
            7348..7367
repeat_region /rpt_family="Simple-repeat"
            7368..7656
repeat_region /rpt_family="AluSx"
            7880..8183
repeat_region /rpt_family="AluSx"
            8484..8778
repeat_region /rpt_family="AluSx"
            9071..9365
repeat_region /rpt_family="AluSx"
            10279..10583
repeat_region /rpt_family="AluSg"
            10609..10864
repeat_region /rpt_family="L2"
            10970..11276
repeat_region /rpt_family="AluJo"
            11276..11594
repeat_region /rpt_family="AluJb"
            11744..11825
repeat_region /rpt_family="L2"
            11764..11987
repeat_region /rpt_family="MIR"
            12989..13307
repeat_region /rpt_family="AluSx"
            13701..13772
exon         /gene="rac1"
            /number=2
            13773..18412
intron       /gene="rac1"
            /number=2
            13954..14235
repeat_region /rpt_family="AluSg"
            14807..14831
repeat_region /rpt_family="Simple-repeat"
            14836..15128
repeat_region /rpt_family="AluSx"
```

```
repeat_region 15408..15715
            /rpt_family="AluSx"
repeat_region 15872..15896
            /rpt_family="AT_rich"
repeat_region 15902..16186
            /rpt_family="AluSx"
repeat_region 16187..16314
            /rpt_family="AluJo"
repeat_region 16413..16585
            /rpt_family="Simple-repeat"
repeat_region 16586..16679
            /rpt_family="FLAM_A"
repeat_region 16707..16810
            /rpt_family="LIM4"
repeat_region 16811..17108
            /rpt_family="AluSg"
repeat_region 17109..17345
            /rpt_family="LIM4"
repeat_region 17377..17686
            /rpt_family="AluSg"
repeat_region 17703..17729
            /rpt_family="Simple-repeat"
exon         18413..18530
            /gene="rac1"
            /number=3
            18531..22803
intron       /gene="rac1"
            /number=3
            19287..19436
repeat_region /rpt_family="MIR"
            21161..21184
repeat_region /rpt_family="Simple-repeat"
            21185..21351
repeat_region /rpt_family="AluSg"
            21365..21403
repeat_region /rpt_family="Alu"
            21364..21365
misc_difference 21364..21365
```

/note="2345 nt insertion sequence in RPO1 11-derived BAC"

Query Match 58.5%; Score 203.6; DB 9; Length 28567;

Best Local Similarity 78.8%; Pred. No. 1.9e-41;

Matches 290; Conservative 0; Mismatches 54; Indels 24; Gaps 3;

```
QY      4 AACATCAGTGTCTTGCCAGAGTACCAACACTGTGATGTGATGATGCCGCCCTCTAGA 63
        |||||||
Db      27782 AACATCAGTGTCTTGCCAGAGTACCGACACTGTGACTGACCAATACTGA-CCCTCTTGA 27840

QY      64 CCTACCCGAGCGGACACATGCTCCGCTA-----CCTTGGGT 102
        |||||
Db      27841 CCTGCCACGCGGACACACGCTCCGTGATGCTGCTTGGCTTATGATGTTCCCTTGGGT 27900

QY      103 CTGTGAGTCTGTGTCAA--GCGCTAGTGTAAACCGCTTCTGTACAACCTAATCACTGG 160
        |||||||
Db      27901 CTGTGAGTCTGTGTAACTGTGATGCTGAGAGATGTTCTGTACAACCTAATCACTGG 27960

QY      161 CAAAGAACAGTGTGGGCTTTCGACCACTAGAACAACTTTTCAATGACAGTTGC 220
        |||||
Db      27961 CGAAGATACAGCGTGGACCTTCAGCACTAGAACAGAAATTTTAAATGACAGTTGC 28020

QY      221 AGAATGTGAGTGTTTTACATTGATCTTGTGCTAATGACAGTATGATTTTGGCA 280
        |||||||
Db      28021 AGAATGTGAGTGTTTTACATTGATCTTGTGCTAATGACAGTATGATTTTGGCA 28080

QY      281 TGTATGACTTAATTAATCTTGAATCATATAAAAAAAAAAAAAAAAAATGCTTTGGAAGTTG 340
        |||||||
Db      28081 TGTATGACTTAATTAATCTTGAATCATATGAGACTGTGAATATACGAGTGTTTTGAAGTTG 28140

QY      341 AAAAAAAAA 348
        |||||
Db      28141 ATGAACAA 28148
```

```

RESULT 11
AC009412      212827 bp   DNA       linear   PRI 07-NOV-2001
LOCUS         Homo sapiens BAC clone RP11-425P5 from 7, complete sequence.
ACCESSION     AC009412
VERSION       AC009412.6  GI:14190769
KEYWORDS      HTG.
SOURCE        Homo sapiens (human)
ORGANISM      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
              Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE     1 (bases 1 to 212827)
AUTHORS       Sulston,J.E. and Waterston,R.
TITLE         Toward a complete human genome sequence
JOURNAL       Genome Res. 8 (11), 1097-1108 (1998)
MEDLINE       99063792
PUBMED        9847074
REFERENCE     2 (bases 1 to 212827)
AUTHORS       Hou,S., Maupin,R., Haakenson,W., Gregory,S. and Belter,E.
TITLE         The sequence of Homo sapiens BAC clone RP11-425P5
JOURNAL       Unpublished
REFERENCE     3 (bases 1 to 212827)
AUTHORS       Waterston,R.H.
TITLE         Direct Submission
JOURNAL       Submitted (21-AUG-1999) Genome Sequencing Center, Washington
              University School of Medicine, 4444 Forest Park Parkway, St. Louis,
              MO 63108, USA
REFERENCE     4 (bases 1 to 212827)
AUTHORS       Waterston,R.H.
TITLE         Direct Submission
JOURNAL       Submitted (23-MAY-2001) Genome Sequencing Center, Washington
              University School of Medicine, 4444 Forest Park Parkway, St. Louis,
              MO 63108, USA
REFERENCE     5 (bases 1 to 212827)
AUTHORS       Waterston,R.
TITLE         Direct Submission
JOURNAL       Submitted (07-NOV-2001) Department of Genetics, Washington
              University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
              On May 23, 2001 this sequence version replaced g1.13431187.
COMMENT       ----- Genome Center
              Center: Washington University Genome Sequencing Center
              Center code: WUGSC
              Web site: http://genome.wustl.edu/gsc
              Contact: sapiens@watson.wustl.edu
              ----- Summary Statistics
              -----
              Center project name: H_NH0425P05

```

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

#### MAPPING INFORMATION:

The sequence of this clone was established as part of a mapping and sequencing collaboration between the NHGRI Chromosome 7 Mapping Project (Eric D. Green, Director), John D. McPherson in the Department of Genetics (Washington University), and the Washington University Genome Sequencing Center. For additional information about the map position of this sequence, see <http://www.nhgri.nih.gov/DIR/CTB/CHR7>, send <mailto:egreen@nhgri.nih.gov>, or see <http://genome.wustl.edu/gsc>

#### SOURCE INFORMATION:

The RPCI-11 human BAC library was made from the blood of one male

#### FEATURES

source

1. 212827

donor, as described by Osogawa,K., Moon,P.Y., Zhao,B., Frengen,E., Tateo,M., Catanesi,J.U. and de Jong,P.J. (1998) An improved approach for construction of bacterial artificial chromosome libraries. Genomics 51:1-8. The clone may be obtained either from Research Genetics, Inc. (<http://www.resgen.com>) or Pieter de Jong and coworkers at the Roswell Park Cancer Institute (<http://bacpac.med.buffalo.edu>)

VECTOR: pBAC3.6

NEIGHBORING SOURCE INFORMATION:

The clone sequenced to the right is CTD-2195F2, 200 bp overlap. Actual start of this clone is at base position 1 of RP11-425P5; actual end is at base position 212633 of RP11-425P5.

Location/Qualifiers

```

repeat_region      1623..1696
                    /organism="Homo sapiens"
                    /mol_type="genomic DNA"
                    /db_xref="taxon:9606"
                    /chromosome="7"
                    /map="7"
                    /clone="RP11-425P5"
                    /clone_1lb="RPCI-11"
                    /rpt_family="MIR"
repeat_region      1698..2006
                    /rpt_family="Alu"
repeat_region      2024..2337
                    /rpt_family="Alu"
repeat_region      3006..3093
                    /rpt_family="L1"
repeat_region      3453..3552
                    /rpt_family="L2"
repeat_region      4924..5046
                    /rpt_family="ERV1"
misc_feature        5318..5693
                    /note="similar to EST BF96284 (NTD:912402607)"
repeat_region      5580..5845
                    /rpt_family="Alu"
repeat_region      5925..6277
                    /rpt_family="MALR"
misc_feature        6842..7041
                    /note="similar to EST BF369507 (NTD:911313152)"
repeat_region      7364..7661
                    /rpt_family="Alu"
repeat_region      8029..8048
                    /rpt_family="(CAAAA)n"
repeat_region      8379..8678
                    /rpt_family="Alu"
misc_feature        8560..8762
                    /note="similar to EST AA699398 (NTD:92702592) z140a03.s1"
repeat_region      8720..9031
                    /rpt_family="Alu"
misc_feature        9083..9560
                    /note="similar to EST AA633603 (NTD:92556817) ae66b02.s1"
repeat_region      9598..9899
                    /rpt_family="Alu"
misc_feature        9871..10098
                    /note="similar to EST AA699398 (NTD:92702592) z140a03.s1"
repeat_region      10746..11044
                    /rpt_family="L1"
repeat_region      11075..11101
                    /rpt_family="AT-rich"
repeat_region      11102..11372
                    /rpt_family="Alu"
repeat_region      11411..11465
                    /rpt_family="GA-rich"
repeat_region      11552..11644
                    /rpt_family="L2"
repeat_region      11685..12193
                    /rpt_family="L1"
repeat_region      12194..12486
                    /rpt_family="Alu"
repeat_region      12487..13003
                    /rpt_family="L1"

```

```

repeat_region      13004..13305
                    /rpt_family="Alu"
repeat_region      13306..13501
                    /rpt_family="L1"
repeat_region      13503..13892
                    /rpt_family="L1"
misc_feature       13893..14224
                    /note="similar to EST BE061102 (NID:g8405752)"
repeat_region      14244..14544
                    /rpt_family="Alu"
repeat_region      14546..14828
                    /rpt_family="Alu"
repeat_region      15091..15189
                    /rpt_family="MIR"
repeat_region      15672..16449
                    /rpt_family="Achoho"
repeat_region      16640..16844
                    /rpt_family="Alu"
repeat_region      16917..16959
                    /rpt_family="Alu"
repeat_region      16960..17287
                    /rpt_family="Achoho"
repeat_region      17288..17572
                    /rpt_family="Achoho"
misc_feature       17610..17938
                    /note="similar to EST BF996427 (NID:g12402750)"
repeat_region      17793..18068
                    /rpt_family="Alu"
repeat_region      18251..18537
                    /rpt_family="Alu"
misc_feature       18485..18884
                    /note="similar to EST BE062347 (NID:g8406997)"
misc_feature       1861..18809
                    /note="similar to EST BG186544 (NID:g13708231)"
repeat_region      18957..19160
                    /rpt_family="L1"
repeat_region      19314..19528
                    /rpt_family="L1"
repeat_region      19819..20135
                    /rpt_family="Alu"
repeat_region      21101..21269
                    /rpt_family="L1"
repeat_region      21882..22203
                    /rpt_family="L1"
misc_feature       22429..22797
                    /note="similar to EST BF993478 (NID:g12399801)"
repeat_region      22717..23007
                    /rpt_family="L1"
misc_feature       22993..23191
                    /note="similar to EST BF767044 (NID:g12115035)"
repeat_region      23167..23471
                    /rpt_family="Alu"
misc_feature       23511..23930
                    /note="similar to EST AW105611 (NID:g6076346) x449g02.x1"
repeat_region      25387..25677
                    /rpt_family="Alu"

Query Match      58.5%: Score 203.6; DB 9; Length 212827;
Best Local Similarity 78.8%; Pred. NO. 2e-41;
Matches 290; Conservative 0; Mismatches 54; Indels 24; Gaps 3;

OY 4 AACATACGCTCTGGCAGATGACCAACTGTCATGTGATGCGCCGCCCTCTAGA 63
    |||||
Db 175810 AACATCAGCTCTGGCAGATGACCGACACTGTGACCAATACTGA-CCCTCTTTA 175868

OY 64 CCTACCCACGCGGACACATGCTTCGGTA-----CCTTGGGT 102
    |||||
Db 175869 CCTCGCCACGCGGACACAGCCTCCTGATGCTTGGCTATGATGTTCCCTTGGGT 175928

OY 103 CTGTGAGTTCTGTCAA--GCCTAGTGTCTAACGCCCTTGTGACCAACTACTCTAGG 160
    |||||
Db 175929 CTGTGAGTTCTGTAAACTGTGCTAGTGTGACGAGATGTCTGTACAACTTAACTCACTGG 175988

```

```

OY 161 CAAGAACAGTGTGGCCCTTTCGACCACTAGAACAAACTTTTTCAAATGACAGTGC 220
    |||||
Db 175989 CGAGAAATACAGCCCTGGACCCCTTCCAGCACTACACAGAAATTTTAAATGACAGTGC 176048

OY 221 AGAATGTGAGAGTGTTTTACATGATGATCTTTGCTAATGACAGTACCATGATGTTTGA 280
    |||||
Db 176049 AGAATGTGAGAGTGTTTTACATGATGATCTTTGCTAATGACAAATACCATGATGTTTGA 176108

OY 281 TGTATGACTTAATTAATCCCTGTAATCATTAATAAAAAAAAAAAAAAAAAATGCTTGAAGTTG 340
    |||||
Db 176109 TGTATGACTTAATTAATCCCTGTAATCATGACGACTGTGAATPACGCGTGTGAGACTTG 176168

OY 341 AAAAAAA 348
    |||||
Db 176169 ATGACAA 176176

RESULT 12
LOCUS AK054993
DEFINITION Homo sapiens cDNA FLJ30431 fis, clone BRACE2008968, highly similar
          to RAS-RELATED C3 BOTULINUM TOXIN SUBSTRATE 1.
ACCESSION AK054993
VERSION AK054993.1 GI:16549633
KEYWORDS oligo capping; fis (full insert sequence).
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE 1
AUTHORS Tashiro,H., Yamazaki,M., Watanabe,K., Kumagai,A., Itakura,S.,
          Fukuzumi,Y., Fujimori,Y., Komiyama,M., Sugiyama,T., Ito,R.,
          Otsuki,T., Sato,H., Wakamatsu,A., Ishii,S., Yamamoto,J., Isono,Y.,
          Kawai-Hio,Y., Saito,K., Nishikawa,T., Kimura,K., Yamaashita,H.,
          Matsuo,K., Nakamura,Y., Sekine,M., Kikuchi,H., Kanda,K.,
          Matsushima,M., Murakawa,K., Kanehori,K., Takahashi-Fujii,A.,
          Oshima,A., Sugiyama,A., Kawakami,B., Suzuki,Y., Sugano,S.,
          Nagahara,K., Masuno,Y., Nagai,K. and Isogai,T.
          NEDO human cDNA sequencing project
          Unpublished
          2 (bases 1 to 2315)
REFERENCE Isogai,T., Otsuki,T. and Sugiyama,T.
AUTHORS Direct Submission
TITLE Submitted (24-OCT-2001) Takao Isogai, Helix Research Institute,
          Genomics Laboratory; 1532-3 Yana, Nisarazu, Chiba 292-0812, Japan
          (E-mail:genomics@helix.co.jp, Tel:81-438-52-3975, Fax:81-438-52-3986)
          NEDO human cDNA sequencing project supported by Ministry of
          Economy, Trade and Industry of Japan; cDNA full insert sequencing:
          Research Association for Biotechnology (RAB); cDNA library
          construction: Helix Research Institute (HRI) (supported by Japan
          Key Technology Center etc.); 5'- & 3'-end one pass sequencing: RAB,
          HRI, and Biotechnology Center, National Institute of Technology and
          Evaluation; clone selection for full insert sequencing: RAB and
          HRI.

FEATURES
source location/Qualifiers
1..2315
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="BRACE2008968"
/tissue_type="cerebellum"
/clone_id="BRACE2"
/note="cloning vector: pME185FL3"

BASE COUNT 603 a 531 c 501 g 680 t

Query Match      55.3%: Score 192.4; DB 9; Length 2315;
Best Local Similarity 81.7%; Pred. NO. 1.2e-38;
Matches 269; Conservative 0; Mismatches 36; Indels 24; Gaps 3;

OY 4 AACATCAGCTCTGGCAGATGACCAACTGTCATGTGATGCGCCGCCCTCTAGA 63
    |||||
Db 1989 AACATCAGCTCTGGCAGATGACCGACACTGTGACCAATACTGA-CCCTCTTTA 2047

```





JOURNAL Genomics 22 (2), 302-312 (1994)  
MEDLINE 95104839  
PubMed 7545969  
REFERENCE 2 (bases 1 to 1232)  
AUTHORS Ishida, Y.  
TITLE Direct Submission  
JOURNAL Submitted (11-NOV-1993) Yoshikazu Ishida, Ikeda Genosphere  
Project/ERATO/JRDC, Tokai University School of Medicine, Bohseidai,  
Isehara, Kanagawa 259-1193, Japan  
(E-mail: shinjien@med.u-tokai.ac.jp, Tel: 81-463-91-5095,  
Fax: 81-463-91-4993)  
On Mar 25, 1994 this sequence version replaced gi:434743.  
Submitted (11-Nov-1993) to DDBJ by:  
Yoshikazu Ishida  
Ikeda Genosphere Project/ERATO/JRDC  
Tokai University School of Medicine  
Bohseidai, Isehara  
Kanagawa 259-11  
Japan  
Phone: 0463-91-4056  
Fax: 0463-91-4110.  
Location/Qualifiers  
1. .1232  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/chromosome="4"  
/clone="PO2ST9"  
/issue\_type="brain striatum"  
1214. .1219  
POLYA\_signal  
POLYA\_site 1232  
BASE COUNT 329 a 240 c 251 g 411 t 1 others  
ORIGIN

Query Match 53.9%; Score 187.4; DB 9; Length 1232;  
Best Local Similarity 80.9%; Pred. No. 2.2e-37;  
Matches 263; Conservative 1; Mismatches 37; Indels 24; Gaps 3;

QY 4 AACATCAGTCTGTGCCAGAGTACCAACACTGTGATGATGCGCCCTCTAGA 63  
|||||  
Db 905 AACATCAGTCTGTGCCAGATTAACGACACTGTGACCAATCTGA-CCCTCTTGA 963  
|||||

QY 64 CCTCACCACGCGGACACATGCTTCGGTGA-----CCTTGGGT 102  
|||||  
Db 964 CCTCGCCACGCGGACACACGCTCCTGTAGTGCCTTGGCTTATGATGTTCTTGGGT 1023  
|||||

QY 103 CTGTGAGTCTGTCAA--GCGCTAGTCTAACGCGTCTGTACAACCTAAGTCACTGG 160  
|||||  
Db 1024 CTGTGAGGTTCTGTAACTGTCTAGTGTGACGATGTTCTGTACAACCTAAGTCACTGG 1083  
|||||

QY 161 CAAGAACACAGTGTGGGCTTGCACACTAGACAACTTTTCAATTGACAGTTGC 220  
|||||  
Db 1084 CGAATAATACAGCGTGGACCTTCAGCACTACACGAATTTTAAATTGACAGTTGC 1143  
|||||

QY 221 AGAATTGTGAGTGTGTTTACATTGATCTTGTCTAATGACAGTATGATGTTTGA 280  
|||||  
Db 1144 AGAATTGTGAGTGTGTTTACATTGATCTTGTCTAATGACATTAATGATGTTTGA 1203  
|||||

QY 281 TGTATGACTTAATAATCCTGAAT 305  
|||||  
Db 1204 TGTATGACTTAATAATCCTTGAA 1228  
|||||

Search completed: August 26, 2003, 02:30:08  
Job time : 1288.94 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 25, 2003, 19:52:32 ; Search time 96.0274 Seconds  
(without alignments)  
9782.674 Million cell updates/sec

Title: US-09-717-321A-17

Perfect score: 348  
Sequence: 1 tgaacacacactgcttgc.....cttggaactgaaaaaaa 348

Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 2552756 seqs, 1349719017 residues

Total number of hits satisfying chosen parameters: 5105512

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 10%  
Listing first 45 summaries

Database :

N\_Geneseq\_19Jun03:\*

- 1: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT:\*
- 2: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT:\*
- 3: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT:\*
- 4: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT:\*
- 5: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT:\*
- 6: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT:\*
- 7: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT:\*
- 8: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT:\*
- 9: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT:\*
- 10: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT:\*
- 11: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT:\*
- 12: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT:\*
- 13: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT:\*
- 14: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT:\*
- 15: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT:\*
- 16: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT:\*
- 17: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT:\*
- 18: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT:\*
- 19: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT:\*
- 20: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:\*
- 21: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:\*
- 22: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:\*
- 23: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT:\*
- 24: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT:\*
- 25: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2003.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	348	100.0	348	22	AAH22399
2	348	100.0	1266	22	AAH22396
3	209.4	60.2	2051	21	AAF18236
4	203.6	58.5	28567	25	ABT17030
5	187.8	54.0	352	22	AAH22400
6	187.4	53.9	1232	24	ABK83623
7	187.4	53.9	1232	24	ABN95134
8	187.4	53.9	1318	23	AAS65569

C	9	163.6	47.0	353	25	ABX46267
C	10	160.6	46.1	447	21	AAA89693
C	11	145.2	41.7	422	21	AAA89694
C	12	123	35.3	123	22	AAH22395
C	13	116.4	33.4	3740	22	AAH22396
C	14	113	32.5	1017	22	AAH22397
C	15	95.4	27.4	201	22	ABX39192
C	16	63.8	18.3	68	16	AAT22394
C	17	60.8	17.5	64	25	ABZ78455
C	18	60.8	17.5	64	25	ABZ09002
C	19	52.4	15.1	162	24	ABK45998
C	20	52.4	15.1	605	21	AAH16478
C	21	52.4	15.1	1022	22	AAH22398
C	22	50.8	14.6	605	24	ABL38065
C	23	50	14.4	50	24	ABZ01605
C	24	43.2	12.4	424	25	ABX46053
C	25	43	12.4	18660	22	AAK79108
C	26	43	12.4	38844	24	AAH40255
C	27	42.2	12.1	640681	24	ABA92787
C	28	42	12.1	15261	22	AAH07498
C	29	41.8	12.0	9817	24	ABL33368
C	30	41.4	11.9	6464	24	ABL32514
C	31	40.8	11.7	5852	12	AAQ11710
C	32	40.4	11.6	501	20	AAH97621
C	33	40.2	11.5	185371	24	ABT10718
C	34	40	11.5	534	22	ABV54457
C	35	39.6	11.4	1494	23	AAK86444
C	36	38.6	11.4	2101	22	AAK86445
C	37	38.6	11.4	17918	24	AAH61419
C	38	39.4	11.3	6059	24	ABL32813
C	39	39.4	11.3	6409	22	AAH46495
C	40	39.2	11.3	277	25	ABX47508
C	41	39.2	11.3	375	25	ABX49849
C	42	39.2	11.3	859	17	AAT32572
C	43	39.2	11.3	2039	17	AAT32569
C	44	39.2	11.3	6621	24	ABL32913
C	45	39.2	11.3	18671	22	AAK90763

#### ALIGNMENTS

RESULT 1	AAH22399	standard; DNA; 348 BP.
ID	AAH22399	
AC	AAH22399;	
XX		
DT	22-AUG-2001	(first entry)
XX		
DE	Human rac1 gene related nucleotide sequence #3.	
XX		
KW	Identification; toxic; hepatotoxic; differential gene expression;	
KW	NSAID; non-steroidal antiinflammatory drug; ds.	
XX		
OS	Homo sapiens.	
XX		
PN	WO200138579-A2.	
XX		
PD	31-MAY-2001.	
XX		
PF	21-NOV-2000; 2000MO-US32049.	
XX		
PR	22-NOV-1999; 99US-0166923.	
PR	18-FEB-2000; 2000US-018531.	
PR	20-NOV-2000; 2000US-0717321.	
PA	(CURA-) CURAGEN CORP.	
XX		
PI	Gould-Rothberg BE, DiIippo VA, Ramsehl TM, Gerweln RW;	
XX		
DR	WPI, 2001-355948/37.	
XX		

Bovine EST associa  
Mouse Exo103 nucle  
Mouse Rab2 nucleot  
Human rac1 genomic  
Human reproductive  
Human rac1 gene re  
Bovine EST associa  
Human gene signatu  
Tumour suppression  
Human oligonucleot  
cDNA encoding colo  
Human colon cancer  
Human rac1 gene re  
Human colon tumour  
Human leukocyte ge  
Bovine EST associa  
Human immune/haema  
Genomic DNA encod  
Buchera sp. genom  
Human reproductive  
Human immune syste  
Dictyostelium plas  
Extended human sec  
Human breast cance  
Human prostate exp  
Human immune/haema  
Human immune regulat  
Human immune syste  
Tumour suppressor  
Bovine EST associa  
Bovine EST associa  
Microspore-specifi  
Rapeseed microspor  
Human immune syste  
Human digestive sy

PT Screening hepatotoxic agent comprises contacting test cell population  
PT expressing RISKMARKER or INJURYMARKER with agent, comparing expression  
PT with reference population and identifying difference in expression  
PT levels -  
PS Disclosure; page 8-9; 76pp; English.  
XX  
CC The present invention describes a method of screening a test agent for  
CC hepatotoxicity. The method comprises: (a) providing a test cell  
CC population comprising a cell capable of expressing one or more nucleic  
CC acid sequences selected from the group consisting of RISKMARKER 1-8  
CC and INJURYMARKER 1-10; (b) contacting the test cell population with a  
CC test agent; (c) measuring expression of one or more of the nucleic  
CC acid sequences in the test cell population; (d) comparing the  
CC expression of the nucleic acid sequence in the test cell population to  
CC the expression of the nucleic acid sequence in a reference cell  
CC population comprising at least one cell whose exposure status to a  
CC hepatotoxic agent is known; and (e) identifying a difference in  
CC expression levels of the RISKMARKER or INJURYMARKER sequences, if  
CC present, in the test cell population and reference cell population.  
CC The method is useful for identifying a hepatotoxic agent. The present  
CC sequence is given in the exemplification of the present invention.  
XX  
SQ Sequence 348 BP; 103 A; 76 C; 67 G; 102 T; 0 other;  
Query Match 100.0%; Score 348; DB 22; Length 348;  
Best Local Similarity 100.0%; Pred. No. 2.5e-77;  
Matches 348; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 TGAACATCATCTCTTGGCAGAGTACCAACTGTCATGTGATGCGCCGCCCTCT 60  
DB 1 TGAACATCATCTCTTGGCAGAGTACCAACTGTCATGTGATGCGCCGCCCTCT 60  
QY 61 AGACCTCACCCAGCGGACACATGCTCCGCTACCTTTGGGCTGTGAGGTTCTGTCAAG 120  
DB 61 AGACCTCACCCAGCGGACACATGCTCCGCTACCTTTGGGCTGTGAGGTTCTGTCAAG 120  
QY 121 CGCTAGTGCCTAACGCCGTTCTGTACAACTTCACTGCGCAAGACACAGTGTGGGCC 180  
DB 121 CGCTAGTGCCTAACGCCGTTCTGTACAACTTCACTGCGCAAGACACAGTGTGGGCC 180  
QY 181 TTTTCGACCACTAGAACAACTTTTTCATTTGACAGTGTGAGATTTGTGGAGTTCTTTTA 240  
DB 181 TTTTCGACCACTAGAACAACTTTTTCATTTGACAGTGTGAGATTTGTGGAGTTCTTTTA 240  
QY 241 CATTCATCTTTTCTAATGACAGTTAGCAGTATGTTTGCATGTATGACTTAATAATCCT 300  
DB 241 CATTCATCTTTTCTAATGACAGTTAGCAGTATGTTTGCATGTATGACTTAATAATCCT 300  
QY 301 TGAATCATTAATAAAAAAAAAAATGCTTTTGGAACTTGAATAAAAA 348  
DB 301 TGAATCATTAATAAAAAAAAAAATGCTTTTGGAACTTGAATAAAAA 348  
RESULT 2  
AAH2396/C  
ID AAH2396 standard; DNA: 1266 BP.  
XX  
AC AAH2396;  
XX  
DT 22-AUG-2001 (first entry)  
XX  
DE Human rac1 contig SEQ ID NO:2.  
XX  
XX Identification; toxic; hepatotoxic; differential gene expression;  
KM NSAIID; non-steroidal antiinflammatory drug; ds.  
XX  
OS Homo sapiens.  
XX  
XX WO200138579-A2.  
XX  
XX 31-MAY-2001.  
XX  
XX

PE 21-NOV-2000; 2000MO-US32049.  
XX  
XX 22-NOV-1999; 99US-0166923.  
PR 18-FEB-2000; 2000US-0183531.  
PR 20-NOV-2000; 2000US-0717321.  
XX  
XX (CURA-) CURAGEN CORP.  
XX  
XX Gould-Rothberg BE, DiIppio VA, Ramsehl TW, Gerwein RW;  
PI  
PI WPI; 2001-355948/37.  
DR  
PT Screening hepatotoxic agent comprises contacting test cell population  
PT expressing RISKMARKER or INJURYMARKER with agent, comparing expression  
PT with reference population and identifying difference in expression  
PT levels -  
PS Disclosure; page 7; 76pp; English.  
XX  
CC The present invention describes a method of screening a test agent for  
CC hepatotoxicity. The method comprises: (a) providing a test cell  
CC population comprising a cell capable of expressing one or more nucleic  
CC acid sequences selected from the group consisting of RISKMARKER 1-8  
CC and INJURYMARKER 1-10; (b) contacting the test cell population with a  
CC test agent; (c) measuring expression of one or more of the nucleic  
CC acid sequences in the test cell population; (d) comparing the  
CC expression of the nucleic acid sequence in the test cell population to  
CC the expression of the nucleic acid sequence in a reference cell  
CC population comprising at least one cell whose exposure status to a  
CC hepatotoxic agent is known; and (e) identifying a difference in  
CC expression levels of the RISKMARKER or INJURYMARKER sequences, if  
CC present, in the test cell population and reference cell population.  
CC The method is useful for identifying a hepatotoxic agent. The present  
CC sequence is given in the exemplification of the present invention.  
XX  
SQ Sequence 1266 BP; 385 A; 258 C; 285 G; 338 T; 0 other;  
Query Match 100.0%; Score 348; DB 22; Length 1266;  
Best Local Similarity 100.0%; Pred. No. 3.5e-77;  
Matches 348; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 TGAACATCATCTCTTGGCAGAGTACCAACTGTCATGTGATGCGCCGCCCTCT 60  
DB 357 TGAACATCATCTCTTGGCAGAGTACCAACTGTCATGTGATGCGCCGCCCTCT 298  
QY 61 AGACCTCACCCAGCGGACACATGCTCCGCTACCTTTGGGCTGTGAGGTTCTGTCAAG 120  
DB 297 AGACCTCACCCAGCGGACACATGCTCCGCTACCTTTGGGCTGTGAGGTTCTGTCAAG 238  
QY 121 CGCTAGTGCCTAACGCCGTTCTGTACAACTTCACTGCGCAAGACACAGTGTGGGCC 180  
DB 121 CGCTAGTGCCTAACGCCGTTCTGTACAACTTCACTGCGCAAGACACAGTGTGGGCC 180  
QY 237 CGCTAGTGCCTAACGCCGTTCTGTACAACTTCACTGCGCAAGACACAGTGTGGGCC 178  
DB 181 TTTTCGACCACTAGAACAACTTTTTCATTTGACAGTGTGAGATTTGTGGAGTTCTTTTA 240  
QY 177 TTTTCGACCACTAGAACAACTTTTTCATTTGACAGTGTGAGATTTGTGGAGTTCTTTTA 118  
QY 241 CATTCATCTTTTCTAATGACAGTTAGCAGTATGTTTGCATGTATGACTTAATAATCCT 300  
DB 117 CATTCATCTTTTCTAATGACAGTTAGCAGTATGTTTGCATGTATGACTTAATAATCCT 300  
QY 301 TGAATCATTAATAAAAAAAAAAATGCTTTTGGAACTTGAATAAAAA 348  
DB 57 TGAATCATTAATAAAAAAAAAAATGCTTTTGGAACTTGAATAAAAA 10  
RESULT 3  
AAFI8236  
ID AAFI8236 standard; DNA: 2051 BP.  
XX  
XX AAFI8236;  
XX  
DT 14-MAR-2001 (first entry)  
XX  
XX

XX Lung cancer associated polynucleotide sequence SEQ ID 255.  
 DE Human; lung cancer associated protein; neuroprotective; cytostatic;  
 KW cardiocactive; immunomodulatory; muscular active; vulnerary;  
 KW gastrointestinal; nephrotoxic; antiinfective; gynecological;  
 KW antibacterial; diagnosis; neural disorder; immune disorder; reproductive;  
 KW proliferative disorder; wound healing; infectious disease; ds.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO20005180-A2.  
 PD 21-SEP-2000.  
 XX  
 PF 08-MAR-2000; 2000WO-US05918.  
 XX  
 PR 12-MAR-1999; 99US-0124270.  
 XX  
 PA (HUMA-) HUMAN GENOME SCI INC.  
 PA (ROSE/) ROSEN C A.  
 XX  
 PI Ruben SM;  
 XX  
 DR WPI: 2000-587514/55.  
 DR P-PSDB: AAB58360.  
 XX  
 PT Lung cancer associated gene sequences, referred to as lung cancer  
 PT antigens, useful for treatment, prevention, and diagnosis of disorders  
 PT such as lung cancer -  
 XX  
 XX Claim 1; Page 716-717; 1425pp; English.  
 PS  
 XX Polynucleotide sequences AAF17982 - AAF18424 encode human lung cancer  
 CC associated proteins represented in AAB58106 - AAB58548. Lung cancer  
 CC associated proteins and polynucleotide sequences, their agonists, and  
 CC antagonists may have neuroprotective; cytostatic; cardioactive;  
 CC immunomodulatory; muscular active general; vulnerary; gastrointestinal  
 CC general; nephrotoxic; antiinfective; gynecological; or antibacterial  
 CC activity. The invention also includes antibodies specific for the  
 CC protein or polynucleotide sequences. The lung cancer associated  
 CC polynucleotide sequences may be used for detection of lung cancer,  
 CC chromosome identification, as chromosome markers, and for numerous other  
 CC diagnostic or research purposes. The proteins may be used to treat  
 CC disorders such as neural, immune, muscular, reproductive,  
 CC gastrointestinal, pulmonary, cardiovascular, renal, and proliferative  
 CC disorders. The proteins may also be used in the treatment of wounds and  
 CC infectious diseases. Polynucleotide sequences AAF18425 - AAF18433 and  
 CC peptide AAB58549 are used in the course of the invention for the  
 CC identification and characterization of the polynucleotide and protein  
 CC sequences.  
 CC  
 XX Sequence 2051 BP; 570 A; 430 C; 433 G; 612 T; 6 other:  
 SQ  
 Query Match 60.2%; Score 209.4; DB 21; Length 2051;  
 Best Local Similarity 82.6%; Pred. No. 1.4e-42;  
 Matches 285; Conservative 0; Mismatches 36; Indels 24; Gaps 3;  
 OY 4 AACATCACTGCTCTCCAGAGTACCAACATGTCATGTGATGAGCGCCGCTCTAGA 63  
 DB 1666 AACATCACTGCTCTCCAGAGTACCAACATGTCATGTGATGAGCGCCGCTCTAGA 1724  
 OY 64 CCTGACCCGACGCGACACATGCTTCCGGTA-----CCTTGGGT 102  
 DB 1725 CCTGACCCGACGCGACACATGCTTCCGGTA-----CCTTGGGT 1784  
 OY 103 CTGTGAGGTCTGTCA--GGCTAGTGTCTAACGCCGTTCTGTACACCTTACCTGCG 160  
 DB 1785 CTGTGAGGTCTGTCA--GGCTAGTGTCTAACGCCGTTCTGTACACCTTACCTGCG 1844  
 OY 161 CAGAGACACAGTGTGGGCGTTTGACACACATGAAACATTTTTCATGTACAGTTGCG 220  
 DB 1845 CAGAGACACAGTGTGGGCGTTTGACACACATGAAACATTTTTCATGTACAGTTGCG 1904

OY 221 AGAATGTGAGTGTCTTTTACATGATCTTTTGTCTAATGACATGACATGATTTGCA 280  
 DB 1905 AGAATGTGAGTGTCTTTTACATGATCTTTTGTCTAATGACATGACATGATTTGCA 1964  
 OY 281 TGTATGACTTAAATTAATCCTTGATCATTAATAAAAAAAAAAAAAA 325  
 DB 1965 TGTATGACTTAAATTAATCCTTGATCATTAATAAAAAAAAAAAAAA 2009  
 RESULT 4  
 ABT17030  
 ID ABT17030 standard; DNA; 28567 BP.  
 XX  
 AC ABT17030;  
 XX  
 DT 03-APR-2003 (first entry)  
 XX  
 XX Human MP21 gene Rac1 SEQ ID No 4.  
 DE  
 XX Cytostatic; p21 pathway modulating agent; cancer; angiogenic; apoptotic;  
 KW cell proliferation disorder; MP21; gene; ds.  
 KW  
 XX Homo sapiens.  
 OS  
 XX WO2003006990-A1.  
 PN  
 XX 23-JAN-2003.  
 PD  
 PF 10-JUL-2002; 2002WO-US21549.  
 XX  
 PR 12-JUL-2001; 2001US-305017P.  
 PR 10-OCT-2001; 2001US-328491P.  
 PR 15-FEB-2002; 2002US-357452P.  
 XX  
 PA (EXEL-) EXELIXIS INC.  
 PI  
 XX Friedman L, Plowman GD, Belvin M, Li D, Funke RP;  
 DR WPI: 2003-221779/21.  
 DR P-PSDB: ABB19756.  
 XX  
 XX Identifying candidate p21 pathway modulator, by contacting an assay  
 PT system having modifiers of p21 polypeptide or gene with a test agent to  
 PT provide a reference activity in system and detecting test agent-biased  
 PT activity -  
 XX  
 XX Examples; Page 56-72; 199pp; English.  
 PS  
 XX The invention relates to a novel method for identifying a candidate p21  
 CC pathway modulating agent. The novel method comprises contacting an assay  
 CC system, comprising a purified MP21 polypeptide (modifier of p21) or  
 CC nucleic acid, with a test agent under conditions, so that but for the  
 CC presence of a test agent, the assay system provides a reference activity  
 CC and detection of test agent-biased activity of the assay system. The  
 CC novel method of the invention is useful for identifying a candidate p21  
 CC pathway modulating agent. The invention also includes a method for  
 CC modulating the p21 pathway of a cell, and a method for diagnosing a  
 CC disease e.g. cancer in a patient. The identified modulators are useful in  
 CC diagnosis, therapy and pharmaceutical development. The modulators are  
 CC useful in a variety of diagnostic and therapeutic applications including  
 CC angiogenic, apoptotic and cell proliferation disorders. This  
 CC polynucleotide sequence represents a gene encoding an MP21 protein of the  
 CC invention.  
 CC  
 XX Sequence 28567 BP; 6762 A; 6358 C; 6796 G; 8651 T; 0 other:  
 SQ  
 Query Match 58.5%; Score 203.6; DB 25; Length 28567;  
 Best Local Similarity 78.8%; Pred. No. 7.5e-41;  
 Matches 290; Conservative 0; Mismatches 54; Indels 24; Gaps 3;  
 OY 4 AACATCACTGCTCTCCAGAGTACCAACATGTCATGTGATGAGCGCCGCTCTAGA 63  
 DB 1666 AACATCACTGCTCTCCAGAGTACCAACATGTCATGTGATGAGCGCCGCTCTAGA 1724



XX Detecting granulocyte activation by detecting differential expression  
PT of genes associated with granulocyte activation, which serves as  
PT diagnostic markers that is useful for monitoring disease states and  
PT drug toxicity -  
XX Claim 1; SEQ ID NO 194; 114pp; English.  
PS  
XX The invention relates to detecting (M1) granulocyte (GC) activation  
CC (GCA), by detecting the level of expression of gene(s) (Gs) identified by  
CC DNA chip analysis as given in the specification, and comparing  
CC the expression level to an expression level in an unactivated  
CC GC, where differential expression of Gs is indicative of GCA.  
CC Also included are modulating (M2) GA by contacting GC with an agent  
CC that alters the expression of at least one gene in Gs; (2) screening (M3)  
CC for an agent capable of modulating GCA or an inflammation (especially  
CC chronic) in a tissue, an allergic response in a subject, exposure of a  
CC subject to a pathogen or sterile inflammatory disease using the  
CC gene expression profile; (3) detecting (M4) an inflammation (especially  
CC chronic) in a tissue, an allergic response in a subject, exposure of a  
CC subject to a pathogen or sterile inflammatory disease, by detecting the  
CC level of expression in a sample of the tissue of gene(s) from Gs, where  
CC the level of expression of the gene is indicative of inflammation;  
CC (4) treating (M5) an inflammation (especially chronic) or in a tissue,  
CC an allergic response in a subject, exposure of a subject to a pathogen  
CC or sterile inflammatory disease, by contacting a tissue having  
CC inflammation with an agent that modulates the expression of gene(s)  
CC from Gs in the tissue. M1 is useful for detecting GCA; M2 is useful for  
CC modulating GA; M3 is useful for screening an agent capable of modulating  
CC GCA preferably in an inflammation in a tissue; M4 is useful for  
CC detecting an inflammation (especially chronic) in a tissue, an allergic  
CC response in a subject, exposure of a subject to a pathogen or sterile  
CC inflammatory disease (e.g. psoriasis, rheumatoid arthritis,  
CC glomerulonephritis, asthma, thrombosis, cardiac reperfusion injury, renal  
CC reperfusion injury, AIDS, adult respiratory distress syndrome,  
CC inflammatory bowel disease, Crohn's disease, ulcerative colitis,  
CC periodontal disease; also bacterial infection, viral infection,  
CC parasitic infection, protozoal infection, fungal infection and M5 is  
CC useful for treating one of the above conditions. The present  
CC sequence represents a gene differentially expressed in granulocytes.  
CC Note: The sequence data for this patent did not form part  
CC of the printed specification, but was obtained in electronic  
CC format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pcl\_sequences.  
XX  
SQ Sequence 1232 BP; 329 A; 240 C; 251 G; 411 T; 1 other:  
Query Match 53.9%; Score 187.4; DB 24; Length 1232;  
Best Local Similarity 80.9%; Pred. No. 3.6e-37;  
Matches 263; Conservative 1; Mismatches 37; Indels 24; Gaps 3;  
OY 4 AACATCACTGCTCTCCAGAGTACCAACATGTCATGATGATGCGCCGCTCTGA 63  
DB 905 AACATCACTGCTCTCCAGATTAACGACATGTCATCTGACCAATACTGA -CCCTCTTGA 963  
OY 64 CCTCACCCAGCGGACACATGCTTCGGTA-----CCTTTGGGT 102  
DB 964 CCTGCGCCACGCGGACACACGCTCTGTAGTCTGCTTGCCTATGATGATGCTTTGGGT 1023  
OY 103 CTGTGAGGTCTGTCA--GGGCTAGTGTAAAGCCGCTTCTGTACACCTACTACACG 160  
DB 1024 CTGTGAGGTCTGTAAACTGTGTAGTGTGACGATGTTGTATCACTTAACCTACACG 1083  
OY 161 CAGAACACAGTGTGGGCTTTGACACACTAGAACACATTTTTCATTTGACAGTTGC 220  
DB 1084 CGAGAAATACACGCGGACGACCTTCAAGCCACTACACAGAAATTTTAAATTGACAGTTGC 1143  
OY 221 AGAATTTGAGAGTGTTTTACATGATCTTTTGTCTAATGACAGTATGATTTTGA 280  
DB 1144 AGAATTTGAGAGTGTTTTACATGATCTTTTGTCTAATGACATATGCAATTTATGTTTGA 1203  
OY 281 TGTATGACTTAAATTAATCTTGAAT 305  
|||||

DB 1204 TGTATGACTTAAATTAATCTTGAAT 1228  
RESULT 7  
ABN95134  
ID ABN95134 standard; DNA; 1232 BP.  
XX  
AC ABN95134;  
XX  
PT 13-AUG-2002 (first entry)  
XX  
DE Gene #1632 used to diagnose liver cancer.  
XX  
KW Gene; liver cancer; ds; hepatocellular carcinoma; hepatotropic;  
KW metastatic liver tumor; cytostatic; expression profile; disease state;  
KW disease progression; drug toxicity; drug efficacy; drug metabolism.  
XX  
OS Homo sapiens.  
XX  
PN WO200229103-A2.  
XX  
PD 11-APR-2002.  
XX  
XX 02-OCT-2001; 2001WO-US30589.  
XX  
PF 02-OCT-2000; 2000US-237054P.  
XX  
PR (GENE-) GENE LOGIC INC.  
XX  
PA Horne D, Alvares C, Peres-Da-Silva S, Vockley JG;  
PI WPI; 2002-426119/45.  
XX  
DR  
XX  
PT Diagnosing and detecting the progression of liver cancer.  
PT hepatocellular carcinoma or metastatic liver tumor in a patient.  
PT Involves detecting the level of expression of two or more genes in a  
PT liver tissue sample -  
XX  
XX Claim 1; SEQ ID NO 1632; 298pp; English.  
XX  
CC The invention relates to a novel method for diagnosing and detecting the  
CC progression of liver cancer, hepatocellular carcinoma or metastatic liver  
CC tumor in a patient, and differentiating metastatic liver cancer from  
CC hepatocellular carcinoma in a patient. Involving detecting the level of  
CC expression of two or more genes represented in ABN93503-ABN97455 in a  
CC tissue sample. The method of the invention has hepatotropic, and  
CC cytostatic activity. The method is useful for diagnosing and detecting  
CC the progression of liver cancer, hepatocellular carcinoma and metastatic  
CC liver carcinoma in a patient. The method is useful for identifying  
CC expression profiles which serve as useful diagnostic markers as well as  
CC markers that can be used to monitor disease states, disease progression,  
CC drug toxicity, drug efficacy and drug metabolism.  
CC Note: The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pcl\_sequences.  
XX  
SQ Sequence 1232 BP; 329 A; 240 C; 251 G; 411 T; 1 other:  
Query Match 53.9%; Score 187.4; DB 24; Length 1232;  
Best Local Similarity 80.9%; Pred. No. 3.6e-37;  
Matches 263; Conservative 1; Mismatches 37; Indels 24; Gaps 3;  
OY 4 AACATCACTGCTCTCCAGAGTACCAACATGTCATGATGATGCGCCGCTCTGA 63  
DB 905 AACATCACTGCTCTCCAGATTAACGACATGTCATCTGACCAATACTGA -CCCTCTTGA 963  
OY 64 CCTCACCCAGCGGACACATGCTTCGGTA-----CCTTTGGGT 102  
DB 964 CCTGCGCCACGCGGACACACGCTCTGTAGTCTGCTTGCCTATGATGATGCTTTGGGT 1023  
OY 103 CTGTGAGGTCTGTCA--GGGCTAGTGTAAAGCCGCTTCTGTACACCTACTACACG 160  
DB 1024 CTGTGAGGTCTGTAAACTGTGTAGTGTGACGATGTTGTATCACTTAACCTACACG 1083  
|||||

OY	161	CAAGAACACAGTGTGGCCCTTTGGACACATAGAACAACTTTTTCATTGACAGTTGC	220
Db	1084	CGAGAAATTCACACGTGGGACCCCTTACGCCATACAGAAATTTTAAATTGACAGTTGC	1143
OY	221	AGAAATGCGAGCTGTTTACATTGCATCTTTGGTAATGCAGTATAGCAGTATGTTTGGCA	280
Db	1144	AGAAATGTGGAGCTGTTTTTACATTGATCTTTTGGCTAATGCATTAGCATTTATGTTTTTGC	1203
OY	281	TGTATGACTTAATAAATCCTTGAAAT	305
Db	1204	TGTATGACTTAATAAATCCTTGAAAT	1228
RESULT 8			
AA565569			
ID	AA565569	standard; cDNA; 1318 BP.	
XX	AA565569;		
AC			
XX	13-FEB-2002	(first entry)	
DT			
XX			
DE	DNA encoding novel human diagnostic protein #1373.		
XX			
KW	Human; chromosome mapping; gene mapping; gene therapy; forensic;		
KW	food supplement; medical imaging; diagnostic; genetic disorder; ss.		
XX			
OS	Homo sapiens.		
PN	WO200175067-A2.		
PD	11-OCT-2001.		
XX			
PE	30-MAR-2001; 2001WO-US08631.		
XX			
PR	31-MAR-2000; 2000US-0540217.		
PR	23-AUG-2000; 2000US-0649167.		
PA	(HYSE-) HYSEQ INC.		
XX			
PI	Dmanac RT, Liu C, Tang YT;		
DR	WPI; 2001-639362/73.		
DR	P-PSDB; ABG01382.		
XX			
PT	New isolated polynucleotide and encoded polypeptides, useful in		
PT	diagnostics, forensics, gene mapping, identification of mutations		
PT	responsible for genetic disorders or other traits and to assess		
PT	biodiversity		
XX			
PS	Claim 1; SEQ ID No 1373; 103pp; English.		
XX			
CC	The invention relates to isolated polynucleotide (I) and		
CC	polypeptide (II) sequences. (I) is useful as hybridisation probes,		
CC	polymerase chain reaction (PCR) primers, oligomers, and for chromosome		
CC	and gene mapping, and in recombinant production of (II). The		
CC	polynucleotides are also used in diagnostics as expressed sequence tags		
CC	for identifying expressed genes. (I) is useful in gene therapy techniques		
CC	to restore normal activity of (II) or to treat disease states involving		
CC	(II). (II) is useful for generating antibodies against it, detecting or		
CC	quantitating a polypeptide in tissue, as molecular weight markers and as		
CC	a food supplement. (II) and its binding partners are useful in medical		
CC	imaging of sites expressing (II). (I) and (II) are useful for treating		
CC	disorders involving aberrant protein expression or biological activity.		
CC	The polypeptide and polynucleotide sequences have applications in		
CC	diagnostics, forensics, gene mapping, identification of mutations		
CC	responsible for genetic disorders or other traits to assess biodiversity		
CC	and to produce other types of data and products dependent on DNA and		
CC	amino acid sequences. AA564197-AA594564 represent novel human		
CC	diagnostic coding sequences of the invention.		
CC	Note: The sequence data for this patent did not appear in the printed		
CC	specification, but was obtained in electronic format directly from WIPO		
CC	at ftp://ipo.int/pub/published_pct_sequences.		

XX	Sequence	1318 BP; 346 A; 268 C; 279 G; 424 T; 1 other;
XX	Query Match	53.9%; Score 187.4; DB 23; Length 1318;
XX	Best Local Similarity	80.9%; Pred. No. 3.7e-37;
XX	Matches	263; Conservative 1; Mismatches 37; Indels 24; Gaps 3;
OY	4	AACATCACTGCTCTTCCGACAGATACACACATCTCATGTGATGATGCCGCCCTCTAGA 63
Db	991	AACATCACTGCTCTTCCGACAGATTAACGACACATGTCATGTGACCAATACTGA-CCCTCTTTA 1049
OY	64	CCTGACCCACCGGACACATCTCTTCCGGTA-----CCTTGGGT 102
Db	1050	CCTGCCCCACCGGACACACACGCCCTCTCTATCGCTTTGGCTATTGATGTTCCTTTGGGT 1109
OY	103	CTGTGAGTCTGTGCA--GCCTAGTGTCTAGCCGCTTGTACACCTTACTACTGG 160
Db	1110	CTGTGAGTCTGTGAAACTGTCTAGTGTGACATGTCTCTGTCACAACTTAACTCACTGG 1169
OY	161	CAAGAACACAGTGTGGGCTTTGGACCACTAGAACAACTTTTCAATTGACAGTTGC 220
Db	1170	CGAGAAATACACGCTGGGACCCCTTCAGCCACTACACAGAAATTTTAAATGACAGTTGC 1229
OY	221	AGAAATTGTGAGTGTTTTACATTGATCTTTGCTAAATGACAGTATGAGATGTTTGA 280
Db	1230	AGAAATTGTGAGTGTTTTACATTGATCTTTGCTAAATGACAAATTATGATTTGTTTGA 1289
OY	281	TGTATGACTTAATAATCCTTGAAAT 305
Db	1290	TGTATGACTTAATAATCCTTGAAAT 1314
RESULT 9		
ABX46267/c		
ID	ABX46267	standard; cDNA: 353 BP.
XX	ABX46267;	
AC		
XX		
DT	21-FEB-2003	(first entry)
XX		
DE	Bovine EST associated with lactation/muscle/fat deposition #11432.	
XX		
KM	Bovine; ss; EST; expressed sequence tag; lactation; LMFD;	
KM	muscle deposition; fat deposition; genome mapping; gene identification;	
XX	gene analysis; cattle breeding.	
OS	Bos Taurus.	
XX		
PN	US2002137139-A1.	
XX		
PD	26-SEP-2002.	
XX		
PF	24-SEP-2001; 2001US-0960352.	
XX		
PR	12-JAN-1999; 99US-115707P.	
PR	11-JAN-2000; 2000US-0480902.	
XX		
PA	(BYAT/) BYATT J C.	
PA	(MATH/) MATHIALAGAN N.	
PA	(TAON/) TAO N.	
XX	(WARR/) WARREN W C.	
XX		
PI	Byatt JC, Mathialagan N, Tao N, Warren WC;	
XX		
DR	WPI; 2003-110599/10.	
XX		
PT	New nucleic acid associated with lactation, and muscle and fat	
PT	deposition, useful for genome mapping, gene identification and	
PT	analysis, cattle breeding, or for genetically improving cattle	
XX		
PS	Claim 2; SEQ ID NO 11432; 245bp; English.	
CC	The invention relates to a purified nucleic acid molecule associated with	







QY 90 GGTACCTTGGCTGTGTGAGGTTCTGTCAAGCGCTAGTGTCTAACGCCGTTCTGTACMACC 149  
DB 123 GGTACCTTGGCTGTGTGAGGTTCTGTCAAGCGCTAGTGTCTAACGCCGTTCTGTACMACC 64  
QY 150 TAACTCAGTGGCAAGACACACTGTTGGCCCTTTGACACACTAGAACAACTTTTTCAA 209  
DB 63 TAATCAGTGGCAAGACACACTGTTGGCCCTTTGACACACTAGAACAACTTTTTCAA 4  
QY 210 TTG 212  
DB 3 TTG 1  
RESULT 13  
AA07296  
ID AA07296 standard; DNA; 3740 BP.  
AC AA07296;  
XX 21-NOV-2001 (first entry)  
DT Human reproductive system related antigen DNA SEQ ID NO: 9984.  
XX Human reproductive system related antigen; reproductive system disorder;  
KM cancer; gene therapy; ds.  
XX Homo sapiens.  
OS WO200155320-A2.  
XX 02-AUG-2001.  
PD 17-JAN-2001; 2001WO-US01339.  
XX 31-JAN-2000; 2000US-0179065.  
PR 04-FEB-2000; 2000US-0180628.  
PR 24-FEB-2000; 2000US-0184664.  
PR 02-MAR-2000; 2000US-0186350.  
PR 16-MAR-2000; 2000US-0189874.  
PR 17-MAR-2000; 2000US-0190076.  
PR 18-APR-2000; 2000US-0198123.  
PR 19-MAY-2000; 2000US-0205515.  
PR 07-JUN-2000; 2000US-0209467.  
PR 28-JUN-2000; 2000US-0214886.  
PR 30-JUN-2000; 2000US-0215135.  
PR 07-JUL-2000; 2000US-0216647.  
PR 07-JUL-2000; 2000US-0216880.  
PR 11-JUL-2000; 2000US-0217487.  
PR 11-JUL-2000; 2000US-0217496.  
PR 14-JUL-2000; 2000US-0218290.  
PR 26-JUL-2000; 2000US-0220963.  
PR 14-AUG-2000; 2000US-0224518.  
PR 14-AUG-2000; 2000US-0224519.  
PR 14-AUG-2000; 2000US-0225213.  
PR 14-AUG-2000; 2000US-0225214.  
PR 14-AUG-2000; 2000US-0225266.  
PR 14-AUG-2000; 2000US-0225267.  
PR 14-AUG-2000; 2000US-0225268.  
PR 14-AUG-2000; 2000US-0225270.  
PR 14-AUG-2000; 2000US-0225447.  
PR 14-AUG-2000; 2000US-0225757.  
PR 14-AUG-2000; 2000US-0225758.  
PR 14-AUG-2000; 2000US-0225759.  
PR 18-AUG-2000; 2000US-0226279.  
PR 22-AUG-2000; 2000US-0226681.  
PR 22-AUG-2000; 2000US-0226868.  
PR 22-AUG-2000; 2000US-0227182.  
PR 23-AUG-2000; 2000US-0227009.  
PR 30-AUG-2000; 2000US-0228924.  
PR 01-SEP-2000; 2000US-0229287.  
PR 01-SEP-2000; 2000US-0229343.  
PR 01-SEP-2000; 2000US-0229344.

PR 01-SEP-2000; 2000US-0229345.  
PR 05-SEP-2000; 2000US-0229509.  
PR 05-SEP-2000; 2000US-0229513.  
PR 06-SEP-2000; 2000US-0230437.  
PR 06-SEP-2000; 2000US-0230438.  
PR 08-SEP-2000; 2000US-0231242.  
PR 08-SEP-2000; 2000US-0231243.  
PR 08-SEP-2000; 2000US-0231244.  
PR 08-SEP-2000; 2000US-0231413.  
PR 08-SEP-2000; 2000US-0231414.  
PR 08-SEP-2000; 2000US-0232080.  
PR 08-SEP-2000; 2000US-0232081.  
PR 12-SEP-2000; 2000US-0231968.  
PR 14-SEP-2000; 2000US-0232397.  
PR 14-SEP-2000; 2000US-0232398.  
PR 14-SEP-2000; 2000US-0232399.  
PR 14-SEP-2000; 2000US-0232400.  
PR 14-SEP-2000; 2000US-0232401.  
PR 14-SEP-2000; 2000US-0233063.  
PR 14-SEP-2000; 2000US-0233064.  
PR 14-SEP-2000; 2000US-0233065.  
PR 21-SEP-2000; 2000US-0234223.  
PR 21-SEP-2000; 2000US-0234274.  
PR 25-SEP-2000; 2000US-0234997.  
PR 25-SEP-2000; 2000US-0234998.  
PR 26-SEP-2000; 2000US-0235484.  
PR 27-SEP-2000; 2000US-0235634.  
PR 27-SEP-2000; 2000US-0235636.  
PR 29-SEP-2000; 2000US-0236327.  
PR 29-SEP-2000; 2000US-0236367.  
PR 29-SEP-2000; 2000US-0236368.  
PR 29-SEP-2000; 2000US-0236369.  
PR 29-SEP-2000; 2000US-0236370.  
PR 02-OCT-2000; 2000US-0236602.  
PR 02-OCT-2000; 2000US-0237037.  
PR 02-OCT-2000; 2000US-0237038.  
PR 02-OCT-2000; 2000US-0237039.  
PR 02-OCT-2000; 2000US-0237040.  
PR 13-OCT-2000; 2000US-0239835.  
PR 13-OCT-2000; 2000US-0239837.  
PR 20-OCT-2000; 2000US-0240960.  
PR 20-OCT-2000; 2000US-0241221.  
PR 20-OCT-2000; 2000US-0241785.  
PR 20-OCT-2000; 2000US-0241786.  
PR 20-OCT-2000; 2000US-0241787.  
PR 20-OCT-2000; 2000US-0241808.  
PR 20-OCT-2000; 2000US-0241809.  
PR 20-OCT-2000; 2000US-0241826.  
PR 01-NOV-2000; 2000US-0244617.  
PR 08-NOV-2000; 2000US-0246474.  
PR 08-NOV-2000; 2000US-0246475.  
PR 08-NOV-2000; 2000US-0246476.  
PR 08-NOV-2000; 2000US-0246477.  
PR 08-NOV-2000; 2000US-0246478.  
PR 08-NOV-2000; 2000US-0246523.  
PR 08-NOV-2000; 2000US-0246524.  
PR 08-NOV-2000; 2000US-0246525.  
PR 08-NOV-2000; 2000US-0246526.  
PR 08-NOV-2000; 2000US-0246527.  
PR 08-NOV-2000; 2000US-0246528.  
PR 08-NOV-2000; 2000US-0246532.  
PR 08-NOV-2000; 2000US-0246609.  
PR 08-NOV-2000; 2000US-0246610.  
PR 08-NOV-2000; 2000US-0246611.  
PR 08-NOV-2000; 2000US-0246613.  
PR 17-NOV-2000; 2000US-0249207.  
PR 17-NOV-2000; 2000US-0249208.  
PR 17-NOV-2000; 2000US-0249209.  
PR 17-NOV-2000; 2000US-0249210.  
PR 17-NOV-2000; 2000US-0249211.  
PR 17-NOV-2000; 2000US-0249212.  
PR 17-NOV-2000; 2000US-0249213.  
PR 17-NOV-2000; 2000US-0249214.



PN US2002137139-A1.  
 XX  
 PD 26-SEP-2002.  
 XX  
 PF 24-SEP-2001; 2001US-0960352.  
 XX  
 PR 12-JAN-1999; 99US-115707P.  
 PR 11-JAN-2000; 2000US-0480902.  
 XX  
 PA (BYAT/) BYAT J C.  
 PA (MATH/) MATHIALAGAN N.  
 PA (TAON/) TAO N.  
 PA (WARR/) WARREN W C.  
 XX  
 PI Byatt JC, Mathialagan N, Tao N, Warren WC;  
 XX  
 DR WPI: 2003-110599/10.  
 XX  
 XX New nucleic acid associated with lactation, and muscle and fat  
 PT deposition, useful for genome mapping, gene identification and  
 PT analysis, cattle breeding, or for genetically improving cattle  
 XX  
 PS Claim 2: SEQ ID No 4357; 245bp; English.

XX The invention relates to a purified nucleic acid molecule associated with  
 CC lactation or muscle and fat deposition (designated LMF), derived  
 CC from cattle, and the LMF nucleic acid can specifically hybridise to a  
 CC second nucleic acid molecule comprising any of 15112 nucleotide  
 CC sequences, appearing as ABX34836-ABX49947, or complements of them.  
 CC Also included are: (1) a transformed cell having a nucleic acid  
 CC comprising an LMF nucleic acid linked to a promoter and a 3' non-  
 CC translated sequence that functions in the cell to cause termination of  
 CC transcription and addition of polyadenylated ribonucleotides to a 3' end  
 CC of the mRNA molecule; and (2) determining a level or pattern of a  
 CC molecule in a bovine cell or tissue comprising: (a) incubating a marker  
 CC nucleic acid (comprising any of the 15112 nucleic acid sequences or its  
 CC complement or fragment) with a complementary nucleic acid molecule  
 CC obtained from the bovine cell or tissue, where hybridisation between the  
 CC marker nucleic acid and the complementary nucleic acid permits the  
 CC detection of the molecule; and (b) detecting the level or pattern of the  
 CC complementary nucleic acid, where the detection of the complementary  
 CC nucleic acid is predictive of the level or pattern of the molecule.  
 CC The LMF nucleic acid is used for determining a level or pattern  
 CC of a molecule in a bovine cell or tissue. It is useful for genome  
 CC mapping, gene identification and analysis, cattle breeding, preparation  
 CC of constructs for use in cattle gene expression, or for genetically  
 CC improving cattle. The present sequence is one of the 15112 bovine  
 CC LMF EST (expressed sequence tag) nucleic acids.  
 CC Note: The present sequence was not shown in the specification but  
 CC was obtained in electronic format from the USPTO web site:  
 CC seqdata.uspto.gov/sequence.html?DocID=20020137139.  
 XX  
 SQ Sequence 201 BP; 67 A; 34 C; 23 G; 77 T; 0 other;

Query Match 27.4%; Score 95.4; DB 25; Length 201;  
 Best Local Similarity 76.5%; Pred. No. 1.9e-14;  
 Matches 117; Conservative 0; Mismatches 36; Indels 0; Gaps 0;

QY 174 TTGGGCTTCGACACATGAGCAACTTTTTCATATGACAGTGCAGATTTGGAGT 233  
 DB 193 TTGAACCTTTTCAAAATAATTAATTTTAAATGACAGTGCAGAAATTTTGAAGT 134  
 QY 234 GTTTTACATGATCTTTGCTAATGACAGTATGTTGTCATGATGATGACTTAAT 293  
 DB 133 GATTTAGATGATGATTTTGAATACGTGATTAAGTATGTTTGAAGTATCCCTAAT 74  
 QY 294 AAATCCTTGAATCAATAAAAAAT 326  
 DB 73 AAATCTTGACCCCTCAGATTAATAAGATTAAT 41

Search completed: August 25, 2003, 23:47:57  
 Job time : 98.0274 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 25, 2003, 19:49:37 ; Search time 723.909 Seconds  
(without alignments)  
11683.734 Million cell updates/sec

Title: US-09-717-321A-17  
Perfect score: 348  
Sequence: 1 tgaacacacactgcttgc.....cttgggaactgaaaaaaa 348

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 22781392 segs, 12152238056 residues  
Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :  
1: em\_estba:\*  
2: em\_esthum:\*  
3: em\_estin:\*  
4: em\_estmu:\*  
5: em\_estov:\*  
6: em\_estp1:\*  
7: em\_estro:\*  
8: em\_hic:\*  
9: gb\_est1:\*  
10: gb\_est2:\*  
11: gb\_hic:\*  
12: gb\_est3:\*  
13: gb\_est4:\*  
14: gb\_est5:\*  
15: em\_estfun:\*  
16: em\_estom:\*  
17: em\_gss\_hum:\*  
18: em\_gss\_hiv:\*  
19: em\_gss\_pln:\*  
20: em\_gss\_vrt:\*  
21: em\_gss\_fun:\*  
22: em\_gss\_mam:\*  
23: em\_gss\_mus:\*  
24: em\_gss\_pro:\*  
25: em\_gss\_fod:\*  
26: em\_gss\_phg:\*  
27: em\_gss\_vr1:\*  
28: gb\_gss1:\*  
29: gb\_gss2:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	324	93.1	429	9	AI059212 UI-R-C1-1
C 2	323.4	92.9	493	10	BF554834 UI-R-E0-C
C 3	323.4	92.9	687	12	BM389059 UI-R-D20-
C 4	321.8	92.5	459	13	BQ780699 UI-R-FF0-

C 5	321.8	92.5	640	13	BQ780117 UI-R-FF0-
C 6	321.4	92.4	477	9	AM520555 UI-R-B0P
C 7	321.4	92.4	489	9	AA859879 UI-R-E0-C
C 8	321.4	92.4	565	12	BI273986 UI-R-CW0-
C 9	321.4	92.4	760	13	BQ191985 UI-R-DRI-
C 10	320.8	92.2	472	13	BQ190077 UI-R-CNI-
C 11	320.8	92.2	615	10	BS380448 UI-R-CBT-
C 12	320.8	92.2	639	10	BE111691 UI-R-BJ1-
C 13	320.8	92.2	705	13	BQ200122 UI-R-EBI-
C 14	318.2	91.4	465	9	AA899757 UI-R-E0-d
C 15	317	91.1	405	10	BE104696 UI-R-BX0-
C 16	317	91.1	412	13	BQ780657 UI-R-FE0-
C 17	317	91.1	517	9	AA859940 UI-R-E0-C
C 18	317	91.1	636	13	BQ780658 UI-R-FE0-
C 19	317	91.1	638	13	BI283790 UI-R-CW0-
C 20	313.8	90.2	469	12	BI276611 UI-R-CX0-
C 21	312.4	89.8	557	12	BG665005 DRABXA06
C 22	310.6	89.3	446	12	BI286743 UI-R-CT0S
C 23	305.4	87.8	463	10	BF548903 UI-R-A0-a
C 24	304.8	87.6	316	9	AI412434 EST240733
C 25	300	86.2	316	9	AI409795 EST238088
C 26	297.2	85.4	407	9	AI009011 EST203462
C 27	297.2	85.4	455	9	AI012196 EST206647
C 28	293.6	84.9	553	9	AA799542 EST189039
C 29	293.6	84.4	427	9	AI178082 EST21741
C 30	292	83.9	431	9	AI410870 EST239163
C 31	285.6	82.1	292	9	AI598992 EST250695
C 32	278	79.9	294	9	AI236740 EST23302
C 33	276.8	79.5	522	14	C06842 C06842 Rat
C 34	273.8	78.7	277	10	BF420446 UI-R-BJ2
C 35	251.2	72.2	277	10	BF420436 UI-R-BJ2
C 36	246.4	70.8	407	10	BF549054 UI-R-A0-a
C 37	243.4	69.9	644	14	CA339407 NISC_1278
C 38	238.4	68.5	256	9	AI234716 EST231278
C 39	227.6	65.4	394	12	BM022196 I60H06.x
C 40	226.6	65.1	322	13	BU757481 UI-1-CFO-
C 41	226.6	65.1	324	13	BU756887 UI-1-CFO-
C 42	226.6	65.1	546	12	BO033391 UI-1-CFO-
C 43	226.6	65.1	565	10	BE996534 UI-M-CG0P
C 44	226	64.9	408	12	BI157798 602923126
C 45	225.8	64.9	404	12	BM502810 1144d06.x

## ALIGNMENTS

RESULT 1  
AI059212/c 429 bp mRNA linear EST 11-FEB-1999  
LOCUS UI-R-C1-1b-e-11-0-UI.s1 UI-R-C1 Rattus norvegicus cDNA clone  
DEFINITION UI-R-C1-1b-e-11-0-UI 3', mRNA sequence.

ACCESSION AI059212  
VERSION AI059212.1 GI:3332989  
KEYWORDS EST.  
SOURCE Rattus norvegicus (Norway rat)  
ORGANISM Rattus norvegicus

REFERENCE 1 (bases 1 to 429)  
AUTHORS Bonaldo,M.F., Lennon,G. and Soares,M.B.  
TITLE Normalization and subtraction: two approaches to facilitate gene discovery  
JOURNAL Genome Res. 6 (9), 791-806 (1996)  
MEDLINE 97044477  
PUBMED 8889548

COMMENT  
Contact: Soares, MB  
Coordinated Laboratory for Computational Genomics  
University of Iowa  
375 Newton Road, 4156 MEBRF, Iowa City, IA 52242, USA  
Tel: 319 335 8250  
Fax: 319 335 9565  
Email: bento-soares@iowa.edu

The sequence tag present in the cDNA between the NotI site and the oligo-dT track served to identify it as a clone from the normalized adult lung library. cDNA Library Preparation: M. Fatima Bonaldo, Ph.D. Clone distribution: clones will be available through Research Genetics

Seq primer: M13 Forward.

## FEATURES

source

```
1..429
Location/Qualifiers
/organism="Rattus norvegicus"
/mol_type="mRNA"
/strain="Sprague-Dawley"
/db_xref="taxon:10116"
/clone="UI-R-C1-1b-e-11-0-UI"
/dev_stage="Adult"
/lab_host="DH10B (Life Technologies)"
/clone_1lb="UI-R-C1"
/notes="Vector: p773D-Pac (Pharmacia) with a modified
polylinker. Site.1: Not I; Site.2: Eco RI; The UI-R-C1
library is a subtracted library derived from the UI-R-C0
library, which is a subtracted library derived from the
UI-R-A1 and UI-R-E1 libraries. The UI-R-A1 library
consisted of a mixture of individually tagged normalized
libraries constructed from rat placenta, adult lung,
brain, liver, kidney, heart, spleen, ovary, and muscle.
The UI-R-E1 library consisted of a mixture of
individually tagged normalized libraries constructed from
8, 12 and 18-day embryo. The tag is a string of 3-5
nucleotides present between the Not I site and the
oligo-dT track which allows identification of the library
of origin of a clone within the mixture. The subtracted
library (UI-R-C1) was constructed as follows: PCR
amplified cDNA inserts from UI-R-C0 clones from which 3'
ESTs had been derived was used as a driver in a
hybridization with the UI-R-C0 library in the form of
single-stranded circles. The remaining single-stranded
circles (subtracted library) was purified by
hydroxyapatite column chromatography, converted to
double-stranded circles and electroporated into DH10B
bacteria (Life Technologies) to generate the UI-R-C1
library. This procedure has been previously described
(Bonaldo, Lennon and Soares, Genome Research 6: 791-806,
1996)"
```

BASE COUNT 130 a 87 c 91 g 121 t

ORIGIN

## Query Match

Best Local Similarity 93.18; Score 324; DB 9; Length 429;  
Matches 333; Conservative 0; Mismatches 15; Indels 0; Gaps 0;

```
1 TGAACATCATCTGCTTGGCAGAGTACCAACACTGTCATGTGATGCGCCCTCT 60
|||||
357 TGAACATCATCTGCTTGGCAGAGTACCAACACTGTCATGTGATGCGCCCTAT 298
|||||
61 AGACCTCACCCAGCGGACACATGCTCCGGTACCTTTGGGTCTGTGAGTCTGTCAAG 120
|||||
297 AGACCTCACCCAGCGGACACATGCTCCGGTACCTTTGGGTCTGTGAGTCTGTCAAG 238
|||||
121 CGCTAGTGTACAGCGCTTGTGTACAACCTTAAGTCTGCTGCTGCTGCTGCTGCTG 180
|||||
237 CGCTAGTGTACAGCGCTTGTGTACAACCTTAAGTCTGCTGCTGCTGCTGCTGCTG 178
|||||
181 TTTGACCACTAGAACAACTTTTTCATTTGACAGTTGACAGTGTGAGAGTCTTTTA 240
|||||
177 TTTGACCACTAGAACAACTTTTTCATTTGACAGTTGACAGTGTGAGAGTCTTTTA 118
|||||
241 CATTTGATCTTTTGTATGATGATGATGATGATGATGATGATGATGATGATGATGAT 300
|||||
117 CATTTGATCTTTTGTATGATGATGATGATGATGATGATGATGATGATGATGATGAT 58
|||||
301 TGAATCATTAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 348
|||||
57 TGAATCATTAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 10
|||||
```

RESULT 2  
BF554834 493 bp mRNA linear EST 12-DEC-2000  
LOCUS  
DEFINITION  
UI-R-E0-cc-c-06-0-UI.r1 UI-R-E0 Rattus norvegicus cDNA clone

ACCESSION  
BF554834  
VERSION  
BF554834.1 GI:11664606  
KEYWORDS  
EST.  
Rattus norvegicus (Norway rat)  
Rattus norvegicus  
Eukaryota; Metazoa; Chordata; Craniala; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;

REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
MEDLINE  
PUBMED  
COMMENT  
discoversy 6 (9), 791-806 (1996)  
Genome Res. 6 (9), 791-806 (1996)  
97044477  
8889548  
Contact: Soares, MB  
Coordinated Laboratory for Computational Genomics  
University of Iowa  
375 Newton Road, 4156 MEBR, Iowa City, IA 52242, USA  
Tel: 319 335 8250  
Fax: 319 335 9565  
Email: Bento-soares@uiowa.edu  
cDNA Library Preparation: M.B. Soares Lab Clone distribution:  
clones will be available through Research Genetics (www.resgen.com)  
This clone is also available through the I.M.A.C.E. Consortium at  
LNLV (info@image.lnl.gov). IMAGE ID: 1770064 The following  
repetitive elements were found in this cDNA sequence: 463-485,  
>AT-richlow-complexity  
Seq primer: M13 Forward.

FEATURES  
source  
Location/Qualifiers

```
1..493
/organism="Rattus norvegicus"
/mol_type="mRNA"
/strain="Sprague-Dawley"
/db_xref="taxon:10116"
/clone="UI-R-E0-cc-c-06-0-UI"
/dev_stage="embryonic"
/lab_host="DH10B (Life Technologies)"
/clone_1lb="UI-R-E0"
/notes="Vector: p773D-Pac (Pharmacia) with a modified
polylinker. Site.1: NotI; Site.2: EcoRI; This library
consists of a mixture of individually tagged normalized
libraries constructed from 8, 12 and 18-day embryo. The
tag is a string of 3-5 nucleotides present between the
Not I site and the oligo-dT track which allows
identification of the library of origin of a clone within
the mixture."
```

BASE COUNT 130 a 109 c 95 g 159 t

ORIGIN

## Query Match

Best Local Similarity 92.98; Score 323.4; DB 10; Length 493;  
Matches 324; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```
1 TGAACATCATCTGCTTGGCAGAGTACCAACACTGTCATGTGATGCGCCCTCT 60
|||||
157 TGAACATCATCTGCTTGGCAGAGTACCAACACTGTCATGTGATGCGCCCTCT 216
|||||
61 AGACCTCACCCAGCGGACACATGCTCCGGTACCTTTGGGTCTGTGAGTCTGTCAAG 120
|||||
217 AGACCTCACCCAGCGGACACATGCTCCGGTACCTTTGGGTCTGTGAGTCTGTCAAG 276
|||||
121 CGCTAGTGTACAGCGCTTGTGTACAACCTTAAGTCTGCTGCTGCTGCTGCTGCTG 180
|||||
277 CGCTAGTGTACAGCGCTTGTGTACAACCTTAAGTCTGCTGCTGCTGCTGCTGCTG 336
|||||
181 TTTGACCACTAGAACAACTTTTTCATTTGACAGTTGACAGTGTGAGAGTCTTTTA 240
|||||
```



Db	337	TTTTGACACACAGAAACAACTTTTTCATATGACAGTTGACGAATGTGGAGTGCTTTT
Oy	241	CATGATCTTTTGGTATGACGTTAGCAGTATGTTTGCATGATGACTTAATAATCTT 300
Db	397	CATTGATCTTTTGGTAAAGCAGTTAGCAGTATGTTTGCATGATGACTTAATAATCTT 456
Oy	301	TGAATCATTAATAAAAAAAAAAAAAA 325
Db	457	TGATCATTAATAAAAAAAAAAAAAA 481
RESULT 3		
BM389059/c		
LOCUS		
DEFINITION		
ACCESSION		
VERSION		
KEYWORDS		
SOURCE		
ORGANISM		
REFERENCE		
AUTHORS		
TITLE		
JOURNAL		
MEDLINE		
PUBMED		
COMMENT		
FEATURES		
SOURCE		

687 bp mRNA linear EST 17-JAN-2002  
 UI-R-D20-cko-j-07-0-UI s1 UI-R-D20 Rattus norvegicus cDNA clone  
 UI-R-D20-cko-j-07-0-UI 3', mRNA sequence.  
 BM389059  
 BM389059.1 GI:18189112  
 EST.  
 Rattus norvegicus (Norway rat)  
 Rattus norvegicus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;  
 Rattus.  
 1 (bases 1 to 687)  
 Bonaldo,M.F., Lennon,G. and Soares,M.B.  
 Normalization and subtraction: two approaches to facilitate gene  
 discovery  
 Genome Res. 6 (9), 791-806 (1996)  
 9704447  
 8889548  
 Contact: Soares, MB  
 Coordinated Laboratory for Computational Genomics  
 University of Iowa  
 375 Newton Road, 4156 MEBRF, Iowa City, IA 52242, USA  
 Tel: 319 335 8250  
 Fax: 319 335 9565  
 Email: bento-soares@uiowa.edu  
 The sequence contained an oligo-dT track that was present in the  
 oligonucleotide that was used to prime the synthesis of first  
 strand cDNA and therefore this may represent a bonafide poly A  
 tail. The sequence tag present in the cDNA between the NotI site  
 and the oligo-dT track served to verify it as a clone from the  
 non-normalized cartilaginous tumor library cDNA library  
 preparation. M.B. Soares lab Clone distribution: clones will be  
 available through Research Genetics (www.resgen.com)  
 Seq primer: M13 Forward  
 POLYA-Yes.

Location/Qualifiers  
 1..687  
 /organism="Rattus norvegicus"  
 /mol\_type="mRNA"  
 /strain="Sprague-Dawley"  
 /db\_xref="taxon:10116"  
 /clone="UI-R-D20-cko-j-07-0-UI"  
 /dex\_stage="adult"  
 /lab\_host="DH10B (Life Technologies)"  
 /clone\_lib="UI-R-D20"  
 /note="Vector: pT73D-Pac (Pharmacia) with a modified  
 polylinker. Site\_1: Not I; Site\_2: Eco RI; UI-R-D20 is a  
 non-normalized Rat cartilaginous tumor library (RCU)  
 constructed in pT73 PAc vector according to the procedure  
 described by Bonaldo, Lennon & Soares (Genome Research  
 Genome 6: 791-806, 1996). The oligonucleotide used to  
 prime first strand synthesis contained the sequence tag  
 CATCTCTGTA between the Not I cloning site and dT18  
 stretch. The Rat cartilaginous tumor tissue was provided  
 by Dr Jeff Stevens at the University of Iowa.  
 TAG\_LIB=UI-R-D20  
 TAG\_TISSUE=cartilaginous tumor  
 TAG\_SEQ=CATCTCTGTA"  
 130 c 163 g 172 t

222 a

ORIGIN	Query Match	92.9%;	Score 323.4;	DB 12;	Length 687;
	Best Local Similarity	99.7%;	Pred. No. 4.3e-37;		
	Matches 324;	Conservative	0;	Mismatches 1;	Indels 0;
					Gaps 0;
QY	1	TGAAACATCACTGTCCTTCCAGAGTACCAACACATGTCAATGTGATGATGCCGCCCTCT	60		
Db	326	TGAAACATCACTGTCCTTCCAGAGTACCAACACATGTCAATGTGATGATGAGCGCCCTCT	267		
QY	61	AGACCTACCCACGCGGACACATGCTTCGCTACCTTTGGGTCTGTAGGTTCTGTCAAG	120		
Db	266	AGACCTACCCACGCGGACACATGCTTCGCTACCTTTGGGTCTGTAGGTTCTGTCAAG	207		
QY	121	CGCTAGTGTCTAAGCGCTTCTGTACAACTTAATCACTGAGGGAAGAACACATGTTGGGCC	180		
Db	206	CGCTAGTGTCTAAGCGCTTCTGTACAACTTAATCACTGAGGGAAGAACACATGTTGGGCC	147		
QY	181	TTTGGACACATAGAACAACTTTTTCAAATGTACAGTGTGCAGAAATGTGGAGCTGTTTTTA	240		
Db	146	TTTGGACACATAGAACAACTTTTTCAAATGTACAGTGTGCAGAAATGTGGAGCTGTTTTTA	87		
QY	241	CATGATCTTTGGCAATGACAGTTAGACGATGTTTTCATGATGATTAATAATCCT	300		
Db	86	CATGATCTTTTGGCAATGACAGTGTAGTGTGATGATGATGACTTAATAATCCT	27		
QY	301	TGAATCATTAATAAAAAAAAAAAAAA 325			
Db	26	TGAATCATTAATAAAAAAAAAAAAAA 2			

FEATURES	SOURCE
<p>REFERENCE</p> <p>AUTHORS</p> <p>TITLE</p> <p>JOURNAL</p> <p>MEDLINE</p> <p>PUBMED</p> <p>COMMENT</p>	<p>RESULT 4</p> <p>LOCUS</p> <p>DEFINITION</p> <p>ACCESSION</p> <p>VERSION</p> <p>KEYWORDS</p> <p>SOURCE</p> <p>ORGANISM</p>
<p>1 (bases 1 to 459)</p> <p>Bonaldo,M.F., Lennon,G. and Soares,M.B.</p> <p>Normalization and subtraction: two approaches to facilitate gene discovery</p> <p>Genome Res. 6 (9), 791-806 (1996)</p> <p>97044477</p> <p>8889348</p> <p>Contact: Soares, MB</p> <p>Coordinated Laboratory for Computational Genomics</p> <p>University of Iowa</p> <p>375 Newton Road , 4156 MEBRF, Iowa City, IA 52242, USA</p> <p>Tel: 319 335 8250</p> <p>Fax: 319 335 9565</p> <p>Email: bento-soares@uiowa.edu</p> <p>Tissue Procurement: Jeff Stevens</p> <p>cDNA library preparation by: Dr. M. Bento Soares, University of Iowa</p> <p>DNA Sequencing by: Dr. M. Bento Soares, University of Iowa</p> <p>Clone Distribution: DISTRIBUTION: Researchers may obtain clones from Research Genetics (www.resgen.com).</p> <p>Seq primer: M13 FORWARD</p> <p>PolyA-yes.</p> <p>Location/Qualifiers</p> <p>1..459</p> <p>/organism="Rattus norvegicus"</p> <p>/mol_type="mRNA"</p> <p>/db_xref="taxon:10116"</p> <p>/clone="UI-R-FF0-cpc-k-22-0-UI"</p> <p>/tissue_type="Mixed tissues"</p>	<p>BO780699/c</p> <p>459 bp</p> <p>mRNA</p> <p>linear</p> <p>EST 26-JUL-2002</p> <p>UI-R-FF0-cpc-k-22-0-UI.g1 UI-R-FF0 Rattus norvegicus cDNA clone</p> <p>UI-R-FF0-cpc-k-22-0-UI 3', mRNA sequence.</p> <p>BQ780699</p> <p>BQ780699.1 GI:21989171</p> <p>EST.</p> <p>Rattus norvegicus (Norway rat)</p> <p>Rattus norvegicus</p> <p>Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;</p> <p>Rattus.</p>

```
/dev_stage="Adult"  
/lab_host="DH10B (Life Technologies) (T1 phage resistant)"  
/clone_lib="UI-R-F0"  
/note="Vector: p773-Pac (Pharmacia) with a modified  
polylinker. Site 1: EcoR I; Site 2: Not I; UI-R-F0 is a  
subtracted cDNA library containing the following tissue(s)  
): Normal cartilage and SR-JWS Tumor line. The  
subtraction was made according to Bonaldo, Lennon and  
Soares, Genome Research, 6:791-806, 1996. The  
oligonucleotide used to prime the synthesis of  
first-strand cDNA contains a library tag sequence that is  
located between the Not I site and the (dri)18 tail. The  
sequence tags for these libraries are: CTATGGAGCG,  
CATCTCTGTA,  
TAG_LIB-UI-R-F0  
TAG_TISSUE-cartilage  
TAG_SBO-CTATGGAGCG"
```

## BASE COUNT

145 a 91 c 103 g 120 t

```
Query Match 92.5%: Score 321.8; DB 13; Length 459;  
Best Local Similarity 99.4%: Pred. No. 9e-37;  
Matches 323; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
```

```
1 TGAACATCATGCTGTGCGAGATACCACTGTCATGATTGCGCCCTCT 60  
325 TGAACATCATGCTGTGCGAGATACCACTGTCATGATTGCGCCCTCT 266  
61 AGACCTCACCCAGCGGACACATGCTCCGTAACCTTTGGGTCTGTGCAAG 120  
265 AGACCTCACCCAGCGGACACATGCTCCGTAACCTTTGGGTCTGTGCAAG 206  
121 CGCTAGTGTAAAGCGCTTGTACAACTCACTGCGCAAGACAGTGTGGCC 180  
205 CGCTAGTGTAAAGCGCTTGTACAACTCACTGCGCAAGACAGTGTGGCC 146  
181 TTTGACCACTAGAACAACTTTTCAATTGACAGTTGAGATTGTGAGTCTT 240  
145 TTTGACCACTAGAACAACTTTTCAATTGACAGTTGAGATTGTGAGTCTT 86  
241 CATTTGATCTTTTGTCTAATGACAGTATGATTTTGCATGATTAATTAATCCT 300  
85 CATTTGATCTTTTGTCTAATGACAGTATGATTTTGCATGATTAATTAATCCT 26  
301 TGAATCATTAATAAAAAAAAAA 325  
25 TGAATCATTAATAAAAAAAAAA 1
```

## RESULT 5

```
LOCUS B0780117/c 640 bp mRNA linear EST 26-JUL-2002  
DEFINITION UI-R-F0-cow-j-10-0-UI.s1 UI-R-F0 Rattus norvegicus cDNA clone  
UI-R-F0-cow-j-10-0-UI 3', mRNA sequence.  
B0780117  
B0780117.1 GI:21988589
```

ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM

EST.  
Rattus norvegicus (Norway rat)  
Rattus norvegicus  
Eukaryota; Metazoa; Chordata; Craniala; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;  
Rattus.

REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
MEDLINE  
PUBMED

1 (pages 1 to 640)  
Bonaldo, M.F., Lennon, G. and Soares, M.B.  
Normalization and subtraction: two approaches to facilitate gene  
discovery  
Genome Res. 6 (9), 791-806 (1996)

COMMENT  
Contact: Soares, MB  
Coordinated Laboratory for Computational Genomics  
University of Iowa  
375 Newton Road, 4156 MEBRF, Iowa City, IA 52242, USA

Tel: 319 335 8250  
Fax: 319 335 9565  
Email: Bento-soares@uiowa.edu  
Tissue Procurement: Jeff Stevens  
cDNA Library Preparation: Dr. M. Bento Soares, University of Iowa  
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa  
Clone Distribution: DISTRIBUTION: Researchers may obtain clones  
from Research Genetics (www.resgen.com).  
Seq primer: M13 FORWARD  
POLYA-Yes.

FEATURES  
source

Location/Qualifiers  
1..640  
/organism="Rattus norvegicus"  
/mol\_type="mRNA"  
/db\_xref="taxon:10116"  
/clone="UI-R-F0-cow-j-10-0-UI"  
/tissue\_type="Mixed tissues"  
/dev\_stage="Adult"  
/lab\_host="DH10B (Life Technologies) (T1 phage resistant)"  
/clone\_lib="UI-R-F0"  
/note="Vector: p773-Pac (Pharmacia) with a modified  
polylinker. Site 1: EcoR I; Site 2: Not I; UI-R-F0 is a  
subtracted cDNA library containing the following tissue(s)  
): Normal cartilage and SR-JWS Tumor line. The  
subtraction was made according to Bonaldo, Lennon and  
Soares, Genome Research, 6:791-806, 1996. The  
oligonucleotide used to prime the synthesis of  
first-strand cDNA contains a library tag sequence that is  
located between the Not I site and the (dri)18 tail. The  
sequence tags for these libraries are: CTATGGAGCG,  
CATCTCTGTA,  
TAG\_LIB-UI-R-F0  
TAG\_TISSUE-rat SRC-JWST tumor line  
TAG\_SBO-CATCTCTGTA"

## BASE COUNT

208 a 121 c 152 g 159 t

```
Query Match 92.5%: Score 321.8; DB 13; Length 640;  
Best Local Similarity 99.4%: Pred. No. 7.5e-37;  
Matches 323; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
```

```
1 TGAACATCATGCTGTGCGAGATACCACTGTCATGATTGCGCCCTCT 60  
326 TGAACATCATGCTGTGCGAGATACCACTGTCATGATTGCGCCCTCT 267  
61 AGACCTCACCCAGCGGACACATGCTCCGTAACCTTTGGGTCTGTGCAAG 120  
266 AGACCTCACCCAGCGGACACATGCTCCGTAACCTTTGGGTCTGTGCAAG 207  
121 CGCTAGTGTAAAGCGCTTGTACAACTCACTGCGCAAGACAGTGTGGCC 180  
206 CGCTAGTGTAAAGCGCTTGTACAACTCACTGCGCAAGACAGTGTGGCC 147  
181 TTTGACCACTAGAACAACTTTTCAATTGACAGTTGAGATTGTGAGTCTT 240  
146 TTTGACCACTAGAACAACTTTTCAATTGACAGTTGAGATTGTGAGTCTT 87  
241 CATTTGATCTTTTGTCTAATGACAGTATGATTTTGCATGATTAATTAATCCT 300  
86 CATTTGATCTTTTGTCTAATGACAGTATGATTTTGCATGATTAATTAATCCT 27  
301 TGAATCATTAATAAAAAAAAAA 325  
26 TGAATCATTAATAAAAAAAAAA 2
```

## RESULT 6

LOCUS AW520555/c 477 bp mRNA linear EST 06-MAR-2000  
DEFINITION UI-R-Bj0p-afx-g-09-0-UI.s1 UI-R-Bj0p Rattus norvegicus cDNA clone  
UI-R-Bj0p-afx-g-09-0-UI 3', mRNA sequence.  
AW520555

```

VERSION      AM520555.1 GI:7162933
KEYWORDS
SOURCE       Rattus norvegicus (Norway rat)
ORGANISM     Rattus norvegicus
              Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
              Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
              Rattus.
REFERENCE    1 (bases 1 to 477)
AUTHORS     Bonaldo,M.F., Lennon,G. and Soares,M.B.
TITLE        Normalization and subtraction: two approaches to facilitate gene
              discovery
JOURNAL      Genome Res. 6 (9), 791-806 (1996)
MEDLINE      97044477
PUBMED       8889548
COMMENT      Contact: Soares, MB
              Coordinated Laboratory for Computational Genomics
              University of Iowa
              375 Newton Road, 4156 MEBRF, Iowa City, IA 52242, USA
              Tel: 319 335 8250
              Fax: 319 335 9565
              Email: bento-soares@uiowa.edu
              The sequence contained an oligo-dT track that was present in the
              oligonucleotide that was used to prime the synthesis of first
              strand cDNA and therefore this may represent a bonafide poly A
              tail. The sequence tag present in the cDNA between the NotI site
              and the oligo-dT track served to identify it as a clone from the
              normalized AV canal at 15 dpc library cDNA library Preparation:
              M.B. Soares Lab Clone distribution: clones will be available
              through Research Genetics (www.resgen.com)
              Seq primer: M13 Forward
              POLYA-Tes.

FEATURES
  source
    1..477
      Location/Qualifiers
        /organism="Rattus norvegicus"
        /mol_type="mRNA"
        /strain="Sprague-Dawley"
        /db_xref="taxon:10116"
        /clone="UI-R-BJ0P-afix-9-09-0-UI"
        /dev_stage="adult"
        /lab_host="DH10B (Life Technologies)"
        /clone_idb="UI-R-BJ0P"
        /note="vector: p773D-Pac (Pharmacia) with a modified
        polylinker. Site_1: Not I; Site_2: Eco RI; The UI-R-BJ0P
        library is a subtracted library derived from the UI-R-A1,
        UI-R-A1, UI-R-A1, UI-R-A1, UI-R-A1, UI-R-A1, and
        UI-R-A1 libraries. These libraries represent tissues from
        rat atrium at 16.5 dpc, ventricle at 16.5 dpc, AV canal
        at 16.5 dpc, atrium at 15 dpc, ventricle at 15 dpc, AV
        canal at 15 dpc, and ventricle at 13 dpc. The tag is a
        string of 5-6 nucleotides present between the Not I site
        and the oligo-dT track. The library was constructed as
        described by Bonaldo, Lennon and Soares, Genome Research
        6: 791-806, 1996.
        TAG_Lib=UI-R-BJ0P
        TAG_TISSUE=AV canal at 15 dpc
        TAG_SEQ=GAAGG"

BASE COUNT      157 a          93 c          103 g          124 t
ORIGIN
Query Match      92.4%; Score 321.4; DB 9; Length 477;
Best Local Similarity 99.7%; Pred. No. 1e-36;
Matches 322; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

1 TGAACATCATCTGTTTGCAGATACCAACATGATTCATGATGCGCCGCT 60
DB TGAACATCATCTGTTTGCAGATACCAACATGATTCATGATGCGCCGCT 264
QY AGACCTCACCCAGCGGACACATGCTCCGATCCTTGAGTCTGTGAAGTCTGCAAG 120
DB AGACCTCACCCAGCGGACACATGCTCCGATCCTTGAGTCTGTGAAGTCTGCAAG 204
QY CGTAGTGTCAAGCGCGTTTGTACACCTAAGTACAGGACACACATGTTGGGCC 180

```

```

DB 203 CGTAGTGTCAAGCGCGTTTGTACACCTAAGTACAGGACACACATGTTGGGCC 144
QY 181 TTTCGACCATCTGAGAACAACTTTTTCATTCAGATGACAGATTTGGAGTGTCTTA 240
DB 143 TTTCGACCATCTGAGAACAACTTTTTCATTCAGATGACAGATTTGGAGTGTCTTA 84
QY 241 CATTGATCTTTTGTCAATGACATGATGATGATGATGATGATGATGATGATGAT 300
DB 83 CATTGATCTTTTGTCAATGACATGATGATGATGATGATGATGATGATGATGAT 24
QY 301 TGAATCATTAATAAAAAAAAAAAAAA 323
DB 23 TGAATCATTAATAATAAAAAAAAAAAAA 1

RESULT 7
AA859879/c 489 bp mRNA linear EST 03-JUL-1999
LOCUS      UI-R-E0-cc-c-06-0-UI.s1 UI-R-E0 Rattus norvegicus cDNA clone
DEFINITION UI-R-E0-cc-c-06-0-UI 3' similar to db|D5274|HUMPO2S79 Human
              randomly sequenced mRNA, mRNA sequence.
ACCESSION  AA859879
VERSION     AA859879.1 GI:4230416
KEYWORDS   EST.
SOURCE      Rattus norvegicus (Norway rat)
ORGANISM    Rattus norvegicus
              Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
              Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
              Rattus.
REFERENCE    1 (bases 1 to 489)
AUTHORS     Bonaldo,M.F., Lennon,G. and Soares,M.B.
TITLE        Normalization and subtraction: two approaches to facilitate gene
              discovery
JOURNAL      Genome Res. 6 (9), 791-806 (1996)
MEDLINE      97044477
PUBMED       8889548
COMMENT      On Mar 10, 1998 this sequence version replaced gi:2949399.
              Contact: Soares, MB
              Coordinated Laboratory for Computational Genomics
              University of Iowa
              375 Newton Road, 4156 MEBRF, Iowa City, IA 52242, USA
              Tel: 319 335 8250
              Fax: 319 335 9565
              Email: bento-soares@uiowa.edu
              The sequence tag present in the cDNA between the NotI site and the
              oligo-dT track served to identify it as a clone from the normalized
              adult 18-day-embryo library. cDNA library Preparation: M. Fatima
              Bonaldo, Ph.D. Clone distribution: clones will be available through
              Research Genetics This clone is also available through the
              I.M.A.G.E. Consortium at LNL (info@image.lnl.gov). IMAGE
              ID=1770064
              Seq primer: M13 Forward
              POLYA-No.

FEATURES
  source
    1..489
      Location/Qualifiers
        /organism="Rattus norvegicus"
        /mol_type="mRNA"
        /strain="Sprague-Dawley"
        /db_xref="taxon:10116"
        /clone="UI-R-E0-cc-c-06-0-UI"
        /dev_stage="embryonic"
        /lab_host="DH10B (Life Technologies)"
        /clone_idb="UI-R-E0"
        /note="vector: p773D-Pac (Pharmacia) with a modified
        polylinker. Site_1: NotI; Site_2: EcoRI; This library
        consists of a mixture of individually tagged normalized
        libraries constructed from 8, 12 and 18-day embryo. The
        tag is a string of 3-5 nucleotides present between the
        Not I site and the oligo-dT track which allows
        identification of the library of origin of a clone within
        the mixture."

BASE COUNT      158 a          96 c          107 g          128 t
ORIGIN

```



Tel: 319 335 8250  
Fax: 319 335 9565

Email: bento-soares@uiowa.edu

The sequence contained an oligo-dT track that was present in the oligonucleotide that was used to prime the synthesis of first strand cDNA and therefore this may represent a bonafide poly A tail. The sequence tag present in the cDNA between the NotI site and the oligo-dT track served to verify it as a clone from the normalized osteoblast library cDNA library preparation: M.B. Soares Lab Clone distribution: clones will be available through Research Genetics (www.resgen.com)  
Seq primer: M13 Forward  
POLYA=yes.

#### FEATURES

source  
Location/Qualifiers  
1..760  
/organism="Rattus norvegicus"  
/mol\_type="mRNA"  
/strain="Sprague-Dawley"  
/db\_xref="taxon:10116"  
/clone="UI-R-DRI-cky-h-11-0-UI"  
/dev\_stage="adult"  
/lab\_host="DH10B (Life Technologies)"  
/clone\_lib="UI-R-DRI"  
/note="Vector: pRT3D-Pac (Pharmacia) with a modified polylinker; Site\_1: Not I; Site\_2: Eco RI; The UI-R-DRI library is a normalized Rat Osteoblast library (nREO) constructed in pRT37 vector according to the procedure described by Bonaldo, Lennon & Soares (Normalization and Subtraction: Two Approaches to Facilitate Gene Discovery. Genome Research 6: 791-806, 1996). The oligonucleotide used to prime first strand synthesis contained the sequence tag AGATATCA between the Not I cloning site and dT18 stretch. The Rat Osteoblast tissue was provided by Lian & Stein of the University of Massachusetts Medical School.  
TAG\_LIB=UI-R-DRI  
TAG\_TISSUE=osteoblast  
TAG\_SEQ=AGATATCA"  
BASE COUNT 244 a 143 c 173 g 198 t 2 others  
ORIGIN

Query Match 92.4%; Score 321.4; DB 13; Length 760;

Best Local Similarity 99.4%; Pred. No. 7.7e-37;

Matches 322; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```

QY 1 TGAACATCATCTGCTTGCAGATGACCAACCTGTCATGTGATGAGCGCCCTCT 60
   |||||||
DB 324 TGAACATCATCTGCTTGCAGATGACCAACCTGTCATGTGATGAGCGCCCTCT 265
   |||||||
QY 61 AGACCTCACCCAGCGGACACATGCTCCGTTGCTGGTCTGTGAGTCTGTCAAG 120
   |||||||
DB 264 AGACCTCACCCAGCGGACACATGCTCCGTTGCTGGTCTGTGAGTCTGTCAAG 205
   |||||||
QY 121 CGCTAGTGTAAAGCCGCTTGTACACCTAACTACAGCAAGAACACAGTGTGGGCC 180
   |||||||
DB 204 CGCTAGTGTAAAGCCGCTTGTACACCTAACTACAGCAAGAACACAGTGTGGGCC 145
   |||||||
QY 181 TTTCAGCAGTGAACAACCTTTTTCATTTGACATGAGCAATTTGGAGTGTTTTA 240
   |||||||
DB 144 TTTCAGCAGTGAACAACCTTTTTCATTTGACATGAGCAATTTGGAGTGTTTTA 85
   |||||||
QY 241 CATTCATCTTTTGTGCTAATGAGATGATGTTTGCATGATGATTAATATCT 300
   |||||||
DB 84 CATTCATCTTTTGTGCTAATGAGATGATGTTTGCATGATGATTAATATCT 25
   |||||||
QY 301 TGAATCATTAATAAAAAAAAAA 324
   |||||||
DB 24 TGAATCATTAATAAAAAAAAAA 1

```

RESULT 10  
BQ190077.c BQ190077 472 bp mRNA linear EST 30-APR-2002  
LOCUS

DEFINITION UI-R-CNI-cjs-h-07-0-UI.s3 UI-R-CNI Rattus norvegicus cDNA clone  
ACCESSION UI-R-CNI-cjs-h-07-0-UI 3', mRNA sequence.  
VERSION BQ190077  
KEYWORDS BQ190077.1 GI:20365628  
SOURCE EST.  
ORGANISM Rattus norvegicus (Norway rat)  
Rattus norvegicus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;  
Rattus.

REFERENCE 1 (bases 1 to 472)  
AUTHORS Bonaldo,M.F., Lennon,G. and Soares,M.B.  
TITLE Normalization and subtraction: two approaches to facilitate gene discovery  
JOURNAL Genome Res. 6 (9), 791-806 (1996)  
MEDLINE 97044477  
PUBMED 8889548  
COMMENT Contact: Soares, MB  
Coordinated Laboratory for Computational Genomics  
University of Iowa  
375 Newton Road, 4156 MEBRF, Iowa City, IA 52242, USA  
Tel: 319 335 8250  
Fax: 319 335 9565  
Email: bento-soares@uiowa.edu

The sequence contained an oligo-dT track that was present in the oligonucleotide that was used to prime the synthesis of first strand cDNA and therefore this may represent a bonafide poly A tail. The sequence tag present in the cDNA between the NotI site and the oligo-dT track served to identify it as a clone from the normalized rat eye library cDNA library preparation: M.B. Soares Lab Clone distribution: clones will be available through Research Genetics (www.resgen.com)  
Seq primer: M13 Forward  
POLYA=yes.

#### FEATURES

source

Location/Qualifiers  
1..472  
/organism="Rattus norvegicus"  
/mol\_type="mRNA"  
/strain="Sprague-Dawley"  
/db\_xref="taxon:10116"  
/clone="UI-R-CNI-cjs-h-07-0-UI"  
/dev\_stage="adult"  
/lab\_host="DH10B (Life Technologies)"  
/clone\_lib="UI-R-CNI"  
/note="Vector: pRT3D-Pac (Pharmacia) with a modified polylinker; Site\_1: Not I; Site\_2: Eco RI; The UI-R-CNI library is a subtracted library derived from the following pool of seven normalized rat libraries: normalized rat seminal vesicles, normalized rat penis, normalized rat bladder, normalized rat cervix, normalized rat brown adipose, normalized rat fundus, and normalized rat salivary gland. It was constructed according to the procedure described by Bonaldo, Lennon & Soares (Genome Research 6: 791-806, 1996). For construction of the CNI library, plasmid DNA from the pool of normalized libraries was electroporated into competent bacteria for the production of single-stranded circular DNA. This was then used as a tracer in a subtractive hybridization with a driver (PCR amplified inserts from a plasmid DNA template preparation) comprising: a) a pool of about 34,000 clones from the Rat Unigene Set corresponding to plates R-5-AN-NN excluding plates R-5-WM and MN. This pool represented 40% of the final driver population. b) a pool of about 29,000 clones from subtracted libraries CNO and CAI corresponding to plates R-CA0-ANV through R-CA0-AXS, R-CA0-AXZ through R-CA0-BAZ, R-CA0-BFE through R-CA0-BHY, R-CA0-BZS, R-CA0-BKE, R-CA0-BKG-H, R-CA0-BKJ-R, R-CA0-BKP through R-CA0-BKS, R-CA0-BKU-V, R-CA0-BLX through R-CA0-BMA, R-CA0-BMC through R-CA0-BME, R-CA0-BNS, R-CA0-BOB through R-CA0-BOU, R-CA0-BPA through R-CA0-BPG, R-CAI-BBA through R-CAI-BDA, R-CAI-BHZ through R-CAI-BUF, R-CAI-BUR, R-CAI-BUT through R-CAI-BKB, R-CAI-BKD, R-CAI-BKF, R-CAI-BKI, R-CAI-BKT, R-CAI-BLF, R-CAI-BLH through

R-CAL-BLN, R-CAL-BLS, R-CAL-BLU-V, R-CAL-BNN, and R-CAL-BLE. The resulting pool represented 20% of the final driver population. c) a pool of about 15,000 clones from non normalized libraries CS0s, CT0s, CU0s, CM0s, CX0s and normalized libraries CS0s, CT0s, CU0s, CM0s, and CX0s corresponding to plates R-CS0s-CBD through R-CS0s-CBO, R-CT0s-CAM through R-CT0s-CAX, R-CU0s-CBP through R-CU0s-CCA, R-CM0s-CCB through R-CM0s-CCM, R-CX0s-CCN through R-CX0s-CCX, R-CS0-BSD, R-CS0-BTD through R-CS0-BTV, R-CS0-BVM, R-CT0-BTW through R-CT0-BUP, R-CT0-BVN, R-CU0-BUQ through R-CU0-BVU, R-CM0-BVY through R-CM0-BMP, R-CM0-BXN through R-CM0-BXO, R-CX0-BMQ through R-CX0-BXM. The resulting pool represented 5% of the final driver population. d) a pool of about 5,000 clones (1,000 from nonnormalized eye library CV0 and 4,000 from normalized eye library CV1) corresponding to plates R-CV0-BRH through R-CV0-BRR, R-CV1-BRS through R-CV1-BSC, R-CV1-BSE through R-CV1-BTC, and R-CV1-BVO through R-CV1-BVU. This pool represented about 5% of the final driver population. e) A pool of about 10,000 clones from subtracted library BS2, BV0 and BV0P (7.9.5 kb cDNA library fraction from rat whole embryo), and BX0 (0.5-7kb cDNA library fraction from rat whole embryo) corresponding to plates R-BS2-BBD through R-BS2-BFB, R-BV0-ANR through R-BV0-ANR, R-BV0P-AOII through R-BV0P-AOX, and R-X0-AQY through R-BX0-ASH. The resulting pool represented 5% of the final driver population. f) a pool of about 7,000 clones from the seven non-normalized libraries that make up the tracer including CY0, CZ0, DA0, DB0, DC0, DD0, and DE0 corresponding to plates R-CY0-BXP through R-CY0-BXK, R-CZ0-BYA through R-CZ0-BYL, R-CZ0-BAB-C, R-DA0-BVJ through R-DA0-BVP, R-DA0-BZD through R-DA0-BZH, R-DB0-BYQ through R-DB0-BZA, R-DC0-BZT through R-DC0-BZQ, R-DC0-CAT through R-DC0-CBA, R-DD0-BZR through R-DD0-CAA, R-DD0-CBB-C, and R-DE0-CAB through R-DE0-CAL. The resulting pool represented about 10% of the final driver population. g) a pool of about 2,000 clones from the pool of normalized libraries, CN0, that makes up the tracer. The corresponding plates are R-CN0-BRW through R-CN0-BLD, R-CN0-BLG, R-CN0-BLE through R-CN0-BLR, R-CN0-BLT, R-CN0-BLM-X, R-CN0-BMP, and R-CN0-BMF through R-CN0-BML. This pool represented 5% of the final driver population. h) a pool of the 28 most abundant clones in the CN0 pool corresponding to the following addresses: bkw-a-09-0-UI, bkw-b-09-0-UI, bkw-b-11-0-UI, bkw-b-10-0-UI, bkw-d-01-0-UI, bkw-d-06-0-UI, bkw-g-08-0-UI, bkw-h-12-0-UI, bky-a-05-0-UI, bky-a-06-0-UI, bky-a-11-0-UI, bky-c-06-0-UI, bky-c-09-0-UI, bky-d-10-0-UI, bky-a-01-0-UI, bly-a-02-0-UI, bly-f-04-0-UI, bly-g-07-0-UI, bly-g-12-0-UI, bly-b-02-0-UI, bly-b-02-0-UI, bly-c-11-0-UI, bly-e-95-0-UI, bly-d-1-08-0-UI, bly-d-02-0-UI, bly-q-h-04-0-UI, bly-r-05-0-UI, bly-f-08-0-UI. This pool represented 5% of the final driver population. i) One abundant CN0 clone (corresponding to the address bky-a-11-0-UI) was digested with Not I and Eco RI and the resulting insert was gel purified. This purified insert was added directly to the driver so that it represented 5% of the final driver population.

TAG-LIB-UI-R-CNI  
TAG-TISSUE-rat eye  
TAG\_SEQ-CAGCC"

Db	264	AGACCTCACCCAGCGGACACATGCTCCCGCTACCTTTGGGTGTGTGAGGTTCTGTCAAG	205
QY	121	CGGTAGTGTCTAAGCGCGCTGTGTCAACCTTACCTGACATGGCAAGAAACACAGTGTGGGCC	180
Db	204	CGCTAGTGTCTAAGCGCGCTGTGTCAACCTTACCTGACATGGCAAGAAACACAGTGTGGGCC	145
QY	181	TTTTCGACCACTAGAACAACTTTTTCATTTGACAGTTTCAGAAATGTGTGAGTGTTTT	240
Db	144	TTTTCGACCACTAGAACAACTTTTTCATTTGACAGTTTCAGAAATGTGTGAGTGTTTT	85
QY	241	CATTCATCTTTCGTAAATGCAGTACAGATGTGTTTGACATGTATGACTTAATTAATCCT	300
Db	84	CATTCATCTTTCGTAAATGCAGTACAGATGTGTTTGACATGTATGACTTAATTAATCCT	25
QY	301	TGAAATCATAAAAAATAAAAA 324	
Db	24	TGAAATCATAAAAAATAAAAA 1	

RESULT	11
LOCUS	BG380448/c
DEFINITION	BG380448 615 bp mRNA linear EST 12-MAR-2001
ACCESSION	U1-R-CTO-btx-d-05-0-U1.s1 U1-R-CrO Rattus norvegicus cDNA clone
VERSION	BG380448 BG380448.1 GI:1304920
KEYWORDS	EST.
SOURCE	Rattus norvegicus (Norway rat)
ORGANISM	Rattus norvegicus

REFERENCE	1 (bases 1 to 615)
AUTHORS	Bonaldi,M.F., Lennon,G. and Soares,M.B.
TITLE	Normalization and subtraction: two approaches to facilitate gene discovery
JOURNAL	Genome Res. 6 (9), 791-806 (1996)
MEDLINE	97044477
PUBMED	8889548
COMMENT	Contact: Soares, MB

COMMENT

Contact: Soares, MB  
Coordinated Laboratory for Computational Genomics  
University of Iowa  
375 Newton Road , 4156 MBRF, Iowa City, IA 52242, USA  
Tel: 319 335 8250  
Fax: 319 335 9565  
Email: bento-soares@uiowa.edu

The sequence contained an oligo-dT track that was present in the oligonucleotide that was used to prime the synthesis of first strand cDNA and therefore this may represent a bonafide poly A tail. The sequence lag present in the cDNA between the NotI site and the oligo-dT track served to verify it as a clone from the normalized rat brain pool library cDNA Library Preparation: M.B. Soares lab clone distribution: clones will be available through Research Genetics ([www.resgen.com](http://www.resgen.com))

Seq primer: M13 Forward  
POLYA-Yes.

	Query Match	Score	320.8	DB 13	Length 472
	Best Local Similarity	99.4%	Pred. No. 1.2e-36		
	Matches	322	Conservative	0	Mismatches 2; Indels 0; Gaps 0;
OY	1	TGAACATCATCTGCTCTGGCAGATACCAACAGCTCATCTGATTTGATGCCGCCCTCTCT	60		
	324	TGAACATCATCTGCTCTTTCACAGATACCAACATCTCATCTGATTTGATGCCGCCCTCTCT	265		
OY	61	AGACCTCACCCAGCGGACACATCTTCCTGGTACCTTTGGCTGTGTAGAGTTTGTCAAG	120		

```

FEATURES
SOURCE
Location/Qualifiers
1. 615
/organism="Rattus norvegicus"
/mol_type="mRNA"
/strain="Sprague-Dawley"
/db_xref="taxon:10116"
/clone="UI-R-CT0-bLx-d-05-0-UI"
/dev_stage="ADULr"
/lab_host="DHI0B (Life Technologies)"
/clone_idb="UI-R-CT0"
/note="Vector: pT73D-Pac (Pharmacia) with a modified
polylinker; Site_1: Not I; Site_2: Eco RI; The UI-R-CT0
library is a normalized library constructed from the
following rat brain tissues: embryonic day 17, embryonic
day 19, embryonic day 21, adult day 1, adult day 12, adult
day 75, adult day 200. For a detailed description of the

```





**TITLE** Normalization and subtraction: two approaches to facilitate gene discovery

**JOURNAL** Genome Res. 6 (9), 791-806 (1996)

**MEDLINE** 97044477

**PUBMED** 8889548

**COMMENT** Contact: Soares, MB  
Coordinated Laboratory for Computational Genomics  
University of Iowa  
375 Newton Road, 4156 MEBRF, Iowa City, IA 52242, USA  
Tel: 319 335 8250  
Fax: 319 335 9565  
Email: bento-soares@uiowa.edu

The sequence contained an oligo-dT track that was present in the oligonucleotide that was used to prime the synthesis of first strand cDNA and therefore this may represent a Bonalde poly A tail. The sequence tag present in the cDNA between the NotI site and the oligo-dT track served to verify it as a clone from the normalized duodenum library cDNA library preparation: M.B. Soares Lab Clone distribution: clones will be available through Research Genetics (www.resgen.com)  
Seq primer: M13 Forward  
POLYA=yes.

**FEATURES**

**source**

1..705

/organism="Rattus norvegicus"

/mol\_type="mRNA"

/strain="Sprague-Dawley"

/db\_xref="taxon:10116"

/clone="UI-R-EB1-clf-1-07-0-UI"

/dev\_stage="adult"

/lab\_host="DH10B (Life Technologies)"

/clone\_lib="UI-R-EB1"

/note="Vector: p773D-Pac (Pharmacia) with a modified polylinker; Site\_1: Not I; Site\_2: Eco RI; UI-R-EB1 is a non-normalized Rat Duodenum library (RDU) constructed in p7737 PAC vector according to the procedure described by Bonaldo, Lennon & Soares (Genome Research Genome 6: 791-806, 1996). The oligonucleotide used to prime first strand synthesis contained the sequence tag TGTGGTTCAT between the Not I cloning site and d118 stretch. The Rat Duodenum tissue was provided by Tom Freeman of the Sanger Center.

TAG LIB-UI-R-EB1

TAG TISSUE=duodenum

TAG SEQ=TGTGGTTCAT

**BASE COUNT** 226 a 136 c 165 g 177 t 1 others

**ORIGIN**

Query Match 92.2%; Score 320.8; DB 13; Length 705;  
Best Local Similarity 99.1%; Pred. No. 9,8e-37;  
Matches 322; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

1 TGAACATCATCTGTTGGCAGATACCAACATGTCATGATGATGCCGCCCTCT 60  
|||||  
326 TGAACATCATCTGTTGGCAGATACCAACATGTCATGATGATGATGCCGCCCT 267  
|||||

61 AGACCTACCCAGCGGACACATGCTCGGTACCTTGGGTCTGAGGTTCTGTAAG 120  
|||||  
266 AGACCTACCCAGCGGACACATGCTCGGTACCTTGGGTCTGAGGTTCTGTAAG 207  
|||||

121 CGCTAGTGTAAAGCCGCTTGTGACAACTTAACCTACCTGCAAGAACAGTGTGGCC 180  
|||||  
206 CGCTAGTGTAAAGCCGCTTGTGACAACTTAACCTACCTGCAAGAACAGTGTGGCC 147  
|||||

181 TTTTCGACCACTAGAACAACTTTTTCATTTGACAGTTGAGAAATGTGAGTGTTTTA 240  
|||||  
146 TTTTCGACCACTAGAACAACTTTTTCATTTGACAGTTGAGAAATGTGAGTGTTTTA 87  
|||||

241 CATTTGATCTTTGCTATGAGTTAGACATGATTTTGCATGATGCTTAATAATCCT 300  
|||||  
86 CATTTGATCTTTGCTATGAGTTAGACATGATTTTGCATGATGCTTAATAATCCT 27  
|||||

301 TGAATCATATAAAAAAAAAAAAAA 325

**Db** 26 TGAATCATATAAAAAAAAAAAAAA 2

**RESULT 14** AA899757/c

**LOCUS** 465 bp mRNA linear EST 05-FEB-1999

**DEFINITION** UI-R-E0-da-b-11-0-UI.s1 UI-R-E0 Rattus norvegicus cDNA clone

**ACCESSION** AA899757

**VERSION** AA899757.1 GI:4232251

**KEYWORDS** EST.

**SOURCE** Rattus norvegicus (Norway rat)

**ORGANISM** Rattus norvegicus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.

**REFERENCE** 1 (bases 1 to 465)  
Bonaldo,M.F., Lennon,G. and Soares,M.B.  
**AUTHORS** Normalization and subtraction: two approaches to facilitate gene discovery

**JOURNAL** Genome Res. 6 (9), 791-806 (1996)

**MEDLINE** 97044477

**PUBMED** 8889548

**COMMENT** On Apr 7, 1998 this sequence version replaced gi:3035111.  
Contact: Soares, MB  
Coordinated Laboratory for Computational Genomics  
University of Iowa  
375 Newton Road, 4156 MEBRF, Iowa City, IA 52242, USA  
Tel: 319 335 8250  
Fax: 319 335 9565  
Email: bento-soares@uiowa.edu  
cDNA Library Preparation: M. Fatima Bonaldo, Ph.D. Clone distribution: clones will be available through Research Genetics  
Seq primer: M13 Forward.

**FEATURES**

**source**

1..465

/organism="Rattus norvegicus"

/mol\_type="mRNA"

/strain="Sprague-Dawley"

/db\_xref="taxon:10116"

/clone="UI-R-E0-da-b-11-0-UI"

/dev\_stage="embryonic"

/lab\_host="DH10B (Life Technologies)"

/clone\_lib="UI-R-E0"

/note="Vector: p773D-Pac (Pharmacia) with a modified polylinker; Site\_1: NotI; Site\_2: EcoRI. This library consists of a mixture of individually tagged normalized libraries constructed from 8, 12 and 18-day embryo. The tag is a string of 3-5 nucleotides present between the Not I site and the oligo-dT track which allows identification of the library of origin of a clone within the mixture."

**BASE COUNT** 155 a 93 c 99 g 118 t

**ORIGIN**

Query Match 91.4%; Score 318.2; DB 9; Length 465;  
Best Local Similarity 99.1%; Pred. No. 2,9e-36;  
Matches 320; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

1 TGAACATCATCTGTTGGCAGATACCAACATGTCATGATGATGCCGCCCTCT 60  
|||||  
323 TGAACATCATCTGTTGGCAGATACCAACATGTCATGATGATGATGCCGCCCTTT 264  
|||||

61 AGACCTACCCAGCGGACACATGCTCGGTACCTTGGGTCTGAGGTTCTGTAAG 120  
|||||  
263 AGACCTACCCAGCGGACACATGCTCGGTACCTTGGGTCTGAGGTTCTGTAAG 204  
|||||

121 CGCTAGTGTAAAGCCGCTTGTGACAACTTAACCTACCTGCAAGAACAGTGTGGCC 180  
|||||  
203 CGCTAGTGTAAAGCCGCTTGTGACAACTTAACCTACCTGCAAGAACAGTGTGGCC 144  
|||||



OY 181 TTTCAGCACCTGAGAACAACTTTTTCATTCAGACAGTTCGAGATGTGGAGCTGTTT 240  
|||||  
DB 143 TTTCAGCACCTGAGAACAACTTTTTCATTCAGACAGTTCGAGATGTGGAGCTGTTT 84  
OY 241 CATTGATCTTTGGTAAATGACAGTTCATGTTTGCATGATGACTTAATAATCCT 300  
|||||  
DB 83 CATTGATCTTTGGTAAATGACAGTTCATGTTTGCATGATGACTTAATAATCCT 24  
OY 301 TGAATCATPAAAAAAAAAAAAA 323  
|||||  
DB 23 TGAATCATPAAAAAAAAAAAAA 1

RESULT 15  
BE104696/c 405 bp mRNA linear EST 13-JUN-2000  
LOCUS BE104696/c  
DEFINITION UI-R-BX0-ars-e-01-0-UI-s1 UI-R-BX0 Rattus norvegicus cDNA clone  
UI-R-BX0-ars-e-01-0-UI 3', mRNA sequence.  
ACCESSION BE104696  
VERSION BE104696.1 GI:8496796  
KEYWORDS EST.  
SOURCE Rattus norvegicus (Norway rat)  
ORGANISM Rattus norvegicus  
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;  
Rattus.  
REFERENCE 1 (bases 1 to 405)  
AUTHORS Bonaldi,M.F., Lennon,G. and Soares,M.B.  
TITLE Normalization and subtraction: two approaches to facilitate gene  
discovery  
JOURNAL Genome Res. 6 (9), 791-806 (1996)  
MEDLINE 97044477  
PUBMED 8889548

COMMENT Contact: Soares, MB  
Coordinated Laboratory for Computational Genomics  
University of Iowa  
375 Newton Road, 4156 MEBRF, Iowa City, IA 52242, USA  
Tel: 319 335 8250  
Fax: 319 335 9565  
Email: bento-soares@iowa.edu  
The sequence contained an oligo-dT track that was present in the  
oligonucleotide that was used to prime the synthesis of first  
strand cDNA and therefore this may represent a bonafide poly A  
tail. cDNA library Preparation: M.B. Soares lab Clone distribution:  
clones will be available through Research Genetics (www.resgen.com)  
The following repetitive elements were found in this cDNA sequence:  
1-22, >AT-rich#Low\_complexity  
Seq primer: M13 Forward  
POLY-A=yes.

FEATURES

source

1..405  
Location/Qualifiers  
/organism="Rattus norvegicus"  
/mol\_type="mRNA"  
/strain="Sprague-Dawley"  
/db\_xref="taxon:10116"  
/clone="UI-R-BX0-ars-e-01-0-UI"  
/dev\_stage="embryonic 13 dpc"  
/lab\_host="DH10B (Life Technologies)"  
/clone\_lib="UI-R-BX0"  
/note="Vector: pT7T3D-Pac (Pharmacia) with a modified  
polylinker; Site\_1: Not I; Site\_2: Eco RI; The UI-R-BX0  
library is derived from 13 dpc whole embryo tissue. For a  
detailed description of the library from which this clone  
was derived, please visit our web site at  
ratseq.eng.uiowa.edu.  
TAG\_SEQ=None found"

BASE COUNT 131 a 76 c 87 g 110 t 1 others  
ORIGIN

Query Match 91.1%; Score 317; DB 10; Length 405;  
Best Local Similarity 98.5%; Pred. No. 4,6e-36;  
Matches 320; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

OY 1 TGAACATCATCTGCTCTTGGCAGAGTACCAACACTGTATGATGTGATGCCGCCCTCT 60  
|||||  
DB 328 TGAACATCATCTGCTCTTGGCAGAGTACCAACACTGTATGATGTGATGCCGCCCTCT 269  
OY 61 AGACCTCAACCCACGGGGGACACATGCTCCGGTACCTTTGGGCTGTGAGTCTGTCAAG 120  
|||||  
DB 268 AGACCTCAACCCACGGGGGACACATGCTCCGGTACCTTTGGGCTGTGAGTCTGTCAAG 209  
OY 121 CCCTAGTGTCAACGGCGTTCTGTACACCTAATCTGACAGTGCAGAAACAGAGTGGGCC 180  
|||||  
DB 208 CCCTAGTGTCAACGGCGTTCTGTACACCTAATCTGACAGTGCAGAAACAGAGTGGGCC 149  
OY 181 TTTCAGCACCTGAGAACAACTTTTTCATTCAGACAGTTCGAGATGTGGAGCTGTTT 240  
|||||  
DB 148 TTTCAGCACCTGAGAACAACTTTTTCATTCAGACAGTTCGAGATGTGGAGCTGTTT 89  
OY 241 CATTGATCTTTGGTAAATGACAGTTCATGTTTGCATGATGACTTAATAATCCT 300  
|||||  
DB 88 CATTGATCTTTGGTAAATGACAGTTCATGTTTGCATGATGACTTAATAATCCT 29  
OY 301 TGAATCATPAAAAAAAAAAAAA 325  
|||||  
DB 28 TGAATCATPAAAAAAAAAAAAA 4

Search completed: August 25, 2003, 23:35:40  
Job time : 727.909 secs

**THIS PAGE BLANK (USPTO)**



```

; TITLE OF INVENTION:  USES THEREOF
; FILE REFERENCE:  CL0000862
; CURRENT APPLICATION NUMBER:  US/09/734,675
; CURRENT FILING DATE:  2000-12-13
; NUMBER OF SEQ ID NOS:  4
; SOFTWARE:  FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 38844
; TYPE:  DNA
; ORGANISM:  Human
; US-09-734-675-3

Query Match      12.4%; Score 43; DB 4; Length 38844;
Best Local Similarity 55.8%; Pred. No. 0.043;
Matches 82; Conservative 0; Mismatches 65; Indels 0; Gaps 0;

QY 201 TTTTTCATTTGACAGTTCGAGATTTGAGTGTTCATTTGATTCCTTTCCTAATGC 260
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 24453 TTTTTCATTTGACAGTTCGAGATTTGAGTGTTCATTTGATTCCTTTCCTAATGC 24512
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||

QY 261 AGTTACAGATGTTTTCATGATGATGATTAATTAATCCTTGATCATATAAAAAA 320
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 24513 TTTATGCAAGAGTTTTCATGATGATGATTAATTAATCCTTGATCATATAAAAAA 24572
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||

QY 321 AAAAATGCTTTTGGAACTTGAAAAA 347
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 24573 AAAGATTATGTGAAATTCGAACTAAA 24599
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||

RESULT 3
; US-07-867-106-2/c
; Sequence 2, Application US/07867106
; Patent No. 5385526
; GENERAL INFORMATION:
; APPLICANT:  Slade, Martin B
; APPLICANT:  Chang, Andy C M
; APPLICANT:  Williams, Keith L
; TITLE OF INVENTION:  Improved Plasmid Vectors for Cellular
; TITLE OF INVENTION:  Slime Moulds of the genus Dictyostelium
; NUMBER OF SEQUENCES:  19
; CORRESPONDENCE ADDRESS:
; ADDRESSEE:  Woodcock Washburn Kurtz Mackiewicz & No. 5389526rls
; STREET:  One Liberty Place 46th Floor
; CITY:  Philadelphia
; STATE:  PA
; COUNTRY:  USA
; ZIP:  19103
; COMPUTER READABLE FORM:
; MEDIUM TYPE:  Floppy disk
; COMPUTER:  IBM PC compatible
; OPERATING SYSTEM:  PC-DOS/MS-DOS
; SOFTWARE:  Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; FILING DATE:  US/07/867,106
; APPLICATION NUMBER:  19920625
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:  AU PJ 7187
; FILING DATE:  02-NOV-1989
; ATTORNEY/AGENT INFORMATION:
; NAME:  Feeney, Joanne Longo
; REGISTRATION NUMBER:  35,134
; REFERENCE/DOCKET NUMBER:  RICE-0002
; TELECOMMUNICATION INFORMATION:
; TELEPHONE:  215-568-3100
; TELEFAX:  215-568-3439
; INFORMATION FOR SEQ ID NO:  2:
; SEQUENCE CHARACTERISTICS:
; LENGTH:  5852 base pairs
; TYPE:  NUCLEIC ACID
; STRANDEDNESS:  single
; TOPOLOGY:  linear
; MOLECULE TYPE:  DNA (genomic)

```

```

; ANTI-SENSE:  NO
; FEATURE:
; NAME/KEY:  CDS
; LOCATION:  2378..5038
; FEATURE:
; NAME/KEY:  CDS
; LOCATION:  2378..5038
; US-07-867-106-2

Query Match      11.7%; Score 40.8; DB 1; Length 5852;
Best Local Similarity 54.7%; Pred. No. 0.094;
Matches 81; Conservative 0; Mismatches 67; Indels 0; Gaps 0;

QY 201 TTTTTCATTTGACAGTTCGAGATTTGAGTGTTCATTTGATTCCTTTCCTAATGC 260
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 5767 TTTAATTAATTTGTAATTTGTAATTTTAAATATGTAATTTGTTGTTGTTTACTT 5708
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||

QY 261 AGTTACAGATGTTTTCATGATGATGATTAATTAATCCTTGATCATATAAAAAA 320
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 5707 ATATTCTATTTTATTATTAATAATTAATTAATTTAATTAATTAATAAAAAA 5648
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||

QY 321 AAAAATGCTTTTGGAACTTGAAAAA 348
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 5647 AAAAATAAATTTAATTAATAAAAAA 5620
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||

```

```

RESULT 4
; US-08-345-756-8/c
; Sequence 8, Application US/08345756
; Patent No. 5633438
; GENERAL INFORMATION:
; APPLICANT:  Baszczyński, Chris
; APPLICANT:  Barbour, Eric
; APPLICANT:  Hattori, Jiro
; APPLICANT:  Mikl, Brian
; TITLE OF INVENTION:  MICROSPORE-SPECIFIC REGULATORY ELEMENT
; NUMBER OF SEQUENCES:  8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE:  Foley & Lardner
; STREET:  3000 K Street, N.W., Suite 500
; CITY:  Washington
; STATE:  D.C.
; COUNTRY:  USA
; ZIP:  20007-5109
; COMPUTER READABLE FORM:
; MEDIUM TYPE:  Floppy disk
; COMPUTER:  IBM PC compatible
; OPERATING SYSTEM:  PC-DOS/MS-DOS
; SOFTWARE:  Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER:  US/08/345,756
; FILING DATE:  22-NOV-1994
; CLASSIFICATION:  800
; ATTORNEY/AGENT INFORMATION:
; NAME:  BENT, STEPHEN A.
; REGISTRATION NUMBER:  29,768
; REFERENCE/DOCKET NUMBER:  33229/236/PIHI
; TELECOMMUNICATION INFORMATION:
; TELEPHONE:  (202)672-5300
; TELEFAX:  (202)672-5399
; TELEX:  904136
; INFORMATION FOR SEQ ID NO:  8:
; SEQUENCE CHARACTERISTICS:
; LENGTH:  859 base pairs
; TYPE:  nucleic acid
; STRANDEDNESS:  single
; TOPOLOGY:  linear
; MOLECULE TYPE:  DNA (genomic)
; US-08-345-756-8

Query Match      11.3%; Score 39.2; DB 1; Length 859;
Best Local Similarity 54.1%; Pred. No. 0.14;
Matches 80; Conservative 0; Mismatches 68; Indels 0; Gaps 0;

```

QY	197	AAACGTTTTCATGAGAGTTCGAGAAATGGGAGAGTGTTCATTCATTCATTTTGGCA	256
Db	240	AAATTTGGTTGGAATTTGTTTCGAATCATGTGGGTTCGATTCAGATTTCGAGTTCGGGA	181
QY	257	ATGCAGTTAGCAGTATGTTTCGATGTATGACTTATATAATCCTGATCATATAAAAAA	316
Db	180	CCCATTTTAATTAATGATATTTTCTTATAAAAAAATTCAAATATACCTTAATGTCGCAAAATCA	121
QY	317	AAAAAAAATGCTTTCGAACCTGAAAA	344
Db	120	AAAAATTAATATATATAAACCTATAAAA	93

RESULT 5  
US-08-625-198-8/c

```

GENERAL INFORMATION:
APPLICANT: Baszczynski, Chris
APPLICANT: Barbour, Eric
APPLICANT: Hattori, Jiro
APPLICANT: Miki, Brian
TITLE OF INVENTION: MICROSPORE-SPECIFIC REGULATORY ELEMENT
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:
ADDRESSEE: Foley & Lardner
STREET: 3000 K Street, N.W., Suite 500
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20007-5109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentL Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/625,198
FILING DATE: 01-APR-1996
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/345,756
FILING DATE: 22-NOV-1994
ATTORNEY/AGENT INFORMATION:
NAME: BENT, STEPHEN A.
REGISTRATION NUMBER: 29,768
REFERENCE/DOCKET NUMBER: 33229/236/PIHI
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202)672-5300
TELEFAX: (202)672-5399
TELEX: 904136
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 859 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-625-198-8

```

Query Match	11.38	Score	39.2	DB	1	length	859		
Best Local Similarity	54.18	Pred. No.	0.14						
Matches	80	Conservative	0	Mismatches	68	Indels	0	Gaps	0

Oy	197	AAACCTTTTCAATGACAGTGGCAGAAATGGGAGTGTTTTAACTATCATCTTTGGTA	256
Db	240	AAITTTGGTTCAATTTGTTCGATCATTGTGGGTTCATTACAGATTCGAGTCCGGTA	181
Oy	257	ATGCAGTTAGCAGTATGTTTTCAGATGATGACTTAATAATCCTTGAAATCAATAAAAAA	316
Db	180	CCCAATTATATATATATTTTCTTAAAAAAATTCAAATATACCTTAATTCGTCMAAATCA	121

```

QY      317 AAAAAAAAAATGCTCTTGGAACTGAAA 344
          ||||| ||| | ||| | |||
Db      120 AAAAAATAAAATATATATAAACATAAAA 93

```

RESULT 6  
US-08-345-756-5/c  
; sequence 5, Application US/08345756  
; patent NO. 5633438  
GENERAL INFORMATION:

APPLICANT: Baszczynski, Chris  
 APPLICANT: Barbour, Eric  
 APPLICANT: Hattori, Jiro  
 APPLICANT: Miki, Brian  
 TITLE OF INVENTION: MICROSPORE-SPECIFIC REGULATORY ELEMENT  
 NUMBER OF SEQUENCES: 8  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Foley & Lardner  
 STREET: 3000 K Street, N.W., Suite 500  
 CITY: Washington  
 STATE: D.C.  
 COUNTRY: USA  
 ZIP: 20007-5109  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: PatentIn Release #1.0, Version #1.30  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/345,756  
 FILING DATE: 22-NOV-1994  
 CLASSIFICATION: 800  
 ATTORNEY/AGENT INFORMATION:  
 NAME: BENT, STEPHEN A.  
 REGISTRATION NUMBER: 29,768  
 REFERENCE/DOCKET NUMBER: 33329/236/PIH1  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (202)672-5300  
 TELEFAX: (202)672-5399  
 TELEX: 904136  
 INFORMATION FOR SEQ ID NO: 5:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 2039 base pairs  
 TYPE: nucleic acid  
 STRANDEDNESS: single  
 TOPOLOGY: linear  
 MOLECULE TYPE: DNA (genomic)  
 US-08-345-756-5

Query Match	11.3%	Score	39.2	DB	1	Length	2039
Best Local Similarity	54.1%	Pred. No.	0.18				
Matches	80	Conservative	0	Mismatches	68	Indels	0
						Gaps	0

Oy	137	AAACCTTTTTCATATGACAGTTCGACGAAATGGGAGATGTTTTTACATATATCTTTTGCA	256
Db	231	AAATTTGGTTTAAATTTGTTTGGAAACATCTGGGGTTCATTCAGATTCGAGTTCGGGTA	172
Oy	257	ATGCAGTAGCAGTATGTTTGCATGATGACCTTAATTAATCCCTAAACATTAATAAAAAA	316
Db	171	CCCATTTTAATATGTAATTTTCTTAAAAAAATTCAAATATACTTAATTCGTAAAAATCA	112
Oy	317	AAAAAAAATGCTTTGGACCTTGAAA	344
Db	111	AAAAATAAATATATATAAAACATATAAAA	84

RESULT 7  
US-08-625-198-5/c  
; Sequence 5, Application US/08625198  
; Patent No. 5755324  
; GENERAL INFORMATION:  
; APPLICANT: Basczynski, Chris  
; APPLICANT: Barbour, Eric

```

APPLICANT: Hattori, Jiro
APPLICANT: Miki, Brian
TITLE OF INVENTION: MICROSPORE-SPECIFIC REGULATORY ELEMENT
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:
ADDRESSEE: Foley & Lardner
STREET: 3000 K Street, N.W., Suite 500
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20007-5109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/625,198
FILING DATE: 01-APR-1996
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/345,756
FILING DATE: 22-NOV-1994
ATTORNEY/AGENT INFORMATION:
NAME: BENT, STEPHEN A.
REGISTRATION NUMBER: 29,768
REFERENCE/DOCKET NUMBER: 35229/236/PTIH
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202)672-5300
TELEFAX: (202)672-5399
TELEX: 904136
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 2039 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
IS-08-625-198-5

```

```

    EARLIER FILING DATE: 1997-06-06
    EARLIER APPLICATION NUMBER: 60/048, 893
    EARLIER FILING DATE: 1997-06-06
    EARLIER APPLICATION NUMBER: 60/048, 900
    EARLIER FILING DATE: 1997-06-06
    EARLIER APPLICATION NUMBER: 60/048, 901
    EARLIER FILING DATE: 1997-06-06
    EARLIER APPLICATION NUMBER: 60/048, 892
    EARLIER FILING DATE: 1997-06-06
    EARLIER APPLICATION NUMBER: 60/048, 915
    EARLIER FILING DATE: 1997-06-06
    EARLIER APPLICATION NUMBER: 60/049, 019
    EARLIER FILING DATE: 1997-06-06
    EARLIER APPLICATION NUMBER: 60/048, 970
    EARLIER FILING DATE: 1997-06-06
    EARLIER APPLICATION NUMBER: 60/048, 972
    EARLIER FILING DATE: 1997-06-06
    EARLIER APPLICATION NUMBER: 60/048, 916
    EARLIER FILING DATE: 1997-06-06
    EARLIER APPLICATION NUMBER: 60/049, 373
    EARLIER FILING DATE: 1997-06-06
    EARLIER APPLICATION NUMBER: 60/048, 875
    EARLIER FILING DATE: 1997-06-06
    EARLIER APPLICATION NUMBER: 60/049, 374
    EARLIER FILING DATE: 1997-06-06
    EARLIER APPLICATION NUMBER: 60/048, 917
    EARLIER FILING DATE: 1997-06-06
    EARLIER APPLICATION NUMBER: 60/048, 949
    EARLIER FILING DATE: 1997-06-06
    EARLIER APPLICATION NUMBER: 60/048, 974
    EARLIER FILING DATE: 1997-06-06
    EARLIER APPLICATION NUMBER: 60/048, 883
    EARLIER FILING DATE: 1997-06-06
    EARLIER APPLICATION NUMBER: 60/048, 897
    EARLIER FILING DATE: 1997-06-06
    EARLIER APPLICATION NUMBER: 60/048, 898
    EARLIER FILING DATE: 1997-06-06
    EARLIER APPLICATION NUMBER: 60/048, 962
    EARLIER FILING DATE: 1997-06-06
    EARLIER APPLICATION NUMBER: 60/048, 963
    EARLIER FILING DATE: 1997-06-06
    EARLIER APPLICATION NUMBER: 60/048, 877
    EARLIER FILING DATE: 1997-06-06
    EARLIER APPLICATION NUMBER: 60/048, 878
    EARLIER FILING DATE: 1997-06-06
    EARLIER APPLICATION NUMBER: 60/070, 923
    EARLIER FILING DATE: 1997-12-18
    EARLIER APPLICATION NUMBER: 60/092, 921
    EARLIER FILING DATE: 1998-07-15
    EARLIER APPLICATION NUMBER: 60/094, 657
    EARLIER FILING DATE: 1998-07-30
    NUMBER OF SEQ ID NOS: 1227
    SOFTWARE: PatentIn Ver. 2.0
    SEQ ID NO 103
    LENGTH: 2218
    TYPE: DNA
    ORGANISM: Homo sapiens
US-09-205-258-103

Query Match      11.1%; Score 38.8; DB 4; Length 2218;
Best Local Similarity 55.1%; Pred. No. 0.24;
Matches 76; Conservative 0; Mismatches 62; Indels 0; Gaps 0;

OY      211 TGAACAGTGCAGAAATTGCGAGCTGTTTTCATATTGACTTTTGCTAATGCAGTAGCACT 270
Db      2061 TGAAGGTATGGTTCCTCTATCTCTTAATTGTAATGTGCTCTTGTGATCARTCATCTTGTTT 2120
OY      271 ATGTTTGCATGTATGCATTAATAATCCCTGAATCATTAATAAAAAAAAAAAAAAATGCT 330
Db      2121 TCCTTTTGACGTATTAATAAACACCTGTTGACAAATATAAAAAAAAAAAAAAAAAAAAA 2180
OY      331 TTGCACTTGAAAAAAA 348

```

```

Db          2181 AAAAAAAAAAAAAAAAAA 2198

RESULT 10
US-09-277-716-17
; Sequence 17, Application US/09277716A
; Patent No. 6232107
; GENERAL INFORMATION:
; APPLICANT: Bryan, Bruce
; APPLICANT: Szent-Gyorgyi, Christopher
; APPLICANT: PROLUME, LTD.
; TITLE OF INVENTION: LUCIFERASES, FLUORESCENT PROTEINS, NUCLEIC ACIDS ENCODING THE
CURRENT APPLICATION NUMBER: US/09/277,716A
CURRENT FILING DATE: 1999-03-26
EARLIER APPLICATION NUMBER: 60/102,939
EARLIER FILING DATE: 1998-10-01
EARLIER APPLICATION NUMBER: 60/089,367
EARLIER FILING DATE: 1998-06-15
EARLIER APPLICATION NUMBER: 60/079,624
EARLIER FILING DATE: 1998-03-27
NUMBER OF SEQ ID NOS: 32
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 17
LENGTH: 1217
TYPE: DNA
ORGANISM: Renilla mulleri
FEATURE:
NAME/KEY: CDS
LOCATION: (31)..(963)
FEATURE:
OTHER INFORMATION: Renilla mulleri Luciferase
US-09-277-716-17

Query Match          10.6%; Score 36.8; DB 3; Length 1217;
Best Local Similarity 63.6%; Pred. No. 0.69;
Matches 56; Conservative 0; Mismatches 32; Indels 0; Gaps 0;

Oy      238 TTACATGACGCTTTGGTATAGCAGTATGTTGCATGTGACTTAATAAT 297
Db      1126 TTCTCTTACGATTATTAATAATACACATCTCTTAATAAATTGTTTAATAAT 1185
Oy      298 CCTGAATCATAAAAAAAAAAAAAAAAA 325
Db      1186 TAATGATTCAGAAAAAAAAAAAAAAAAA 1213

RESULT 11
US-09-609-161B-17
; Sequence 17, Application US/09609161B
; Patent No. 6436682
; GENERAL INFORMATION:
; APPLICANT: Bryan, Bruce
; APPLICANT: Szent-Gyorgyi, Christopher
; APPLICANT: PROLUME, LTD.
; TITLE OF INVENTION: LUCIFERASES, FLUORESCENT PROTEINS, NUCLEIC ACIDS ENCODING THE LUC
TITLE OF INVENTION: AND FLUORESCENT PROTEIN AND THE USE THEREOF IN DIAGNOSTICS, HIG
FILE REFERENCE: 24729-1121B
CURRENT APPLICATION NUMBER: US/09/609,161B
CURRENT FILING DATE: 2000-06-30
PRIOR APPLICATION NUMBER: 09/277,716
PRIOR FILING DATE: 1999-03-26
PRIOR APPLICATION NUMBER: 60/102,939
PRIOR FILING DATE: 1998-10-01
PRIOR APPLICATION NUMBER: 60/089,367
PRIOR FILING DATE: 1998-06-15
PRIOR APPLICATION NUMBER: 60/079,624
PRIOR FILING DATE: 1998-03-27
NUMBER OF SEQ ID NOS: 32
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 17
LENGTH: 1217
TYPE: DNA

```

```

      : ORGANISM: Renilla mulleri
      : FEATURE:
      : NAME/KEY: CDS
      : LOCATION: (31)..(963)
      : OTHER INFORMATION: Renilla mulleri Luciferase
      : US-09-609-161B-17

Query Match          10.6%; Score 36.8; DB 4; Length 1217;
Best Local Similarity 63.6%; Pred. No. 0.69;
Matches 56; Conservative 0; Mismatches 32; Indels 0; Gaps 0;

Qy      238 TTACATGTCATTCTTGTCTAATGCAGTTAGCAGTAATGTTTTGCATGTAAGCTTAATAAT 297
Db      1126 TTTCTCTTAGCATTAATTAATATCACACTTCTATGAATAAACTTGTTTAAATAAT 1185
      :
Oy      298 CCTGAATCATATAAAAAAAAAAAAAAAA 325
Db      1186 TAATGATTCAGAAAAAATAAAAAAAAAA 1213
      :
RESULT 12
US-08-745-995A-31
: Sequence 31, Application US/08745995A
: Patent No. 6372887
: GENERAL INFORMATION:
: APPLICANT: Silver, Gary M.
: APPLICANT: Wismeski, Nancy
: TITLE OF INVENTION: No. 6372887el Serine Protease Inhibitor
: TITLE OF INVENTION: Nucleic Acid Molecules, Proteins
: NUMBER OF SEQUENCES: 41
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Heska Corporation
: STREET: 1825 Sharp Point Drive
: CITY: Fort Collins
: STATE: Colorado
: COUNTRY: USA
: ZIP: 80525
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: Windows 95
: SOFTWARE: Wordperfect for Windows, Version 7.0
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/745, 995A
: FILING DATE: 07-NOV-1996
: CLASSIFICATION: 536
: ATTORNEY/AGENT INFORMATION:
: NAME: Verser, Carol Talkington
: REGISTRATION NUMBER: 37,459
: REFERENCE/DOCKET NUMBER: RC-2
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 970/493-7272
: TELEFAX: 970/484-9505
: INFORMATION FOR SEQ ID NO: 31:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1454 nucleotides
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: cDNA
: FEATURE:
: NAME/KEY: CDS
: LOCATION: 20...1210
: US-08-745-995A-31

Query Match          10.5%; Score 36.4; DB 4; Length 1454;
Best Local Similarity 58.2%; Pred. No. 0.93;
Matches 64; Conservative 0; Mismatches 46; Indels 0; Gaps 0;

Oy      216 GTTCCAGCAATGTTGGAGTGTITTTTACATGTGATCTTTTGCTAATGCAGTACAGTATGTT 275
Db      1345 GTACACAAAAGTCTTTGTGTTTGGTTTTCACCTTTTATGATGTAATCACTTATTAAT 1404

```





FILED DATE: 07-NOV-1996  
CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:  
NAME: Verser, Carol Talkington  
REGISTRATION NUMBER: 37,459  
REFERENCE/DOCKET NUMBER: FC-2  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 970/493-7272  
TELEFAX: 970/484-9505  
INFORMATION FOR SEQ ID NO: 33:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1454 nucleotides  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: CDNA  
US-09-005-352-33

Query Match 10.58; Score 36.4; DB 4; Length 1454;  
Best Local Similarity 58.28; Pred. No. 0.93;  
Matches 64; Conservative 0; Mismatches 46; Indels 0; Gaps 0;  
QY 216 GTTGCAGAAATTGCGAGTGTGTTTACATTGATCTTTTGGCTAATGACAGTTAGCAGTATGTT 275  
DB 110 GTAGAGGAAAAAGTTTGTAGTTTCACCTTTTATGATGATCATCACCTATATAT 51  
QY 276 TTGCATGTATGACTTAATAAATCCCTGATCATTAATAAAAAAAAAAAAAAAAA 325  
DB 50 GTGTAGTTTATGTAAATAATGTAAATGTGAAAAAATAAAAAAAAAAAAAAAAA 1

Search completed: August 26, 2003, 02:33:16  
Job time: 26.5279 secs









```

1 CURRENT FILING DATE: 2001-12-26
2 NUMBER OF SEQ ID NOS: 1896
3 SOFTWARE: FASTSEQ for Windows Version 4.0
4 SEQ ID NO 1549
5 LENGTH: 162
6 TYPE: DNA
7 ORGANISM: Homo sapiens
8 US-10-033-528-1549

```

Query Match	15.1%;	Score 52.4;	DB 13;	Length 162;
Best Local Similarity	81.1%;	Pred. No. 0.00028;		
Matches 73; Conservative	0;	Mismatches 16;	Indels 1;	Gaps 1.

09 4 AACACACACTGCTTGGCAGAGTACCAACACTGTATGTGATTGATGCGCCGCCCTCTAGA 633  
|||||  
39 AACATCACTGCTTGGCAGATTACCGACACTGTCACTTGACCAATCACTGA-CCCTCTTTTA 977

**Dy**           64 CCTCACCACGCGGCACATGCTTCGGTA 93  
               |||||  
**Db**           98 CCTGCCACGCGGCACACAGCCCTCTGTGA 127

RESULT 11  
US-10-066-543-3022/c  
: Sequence 3022. Application HS/10066543

```

;
; PUBLICATION NO.: 032003008/01061
;
; GENERAL INFORMATION:
;
; APPLICANT: Jiang, Yugu
;
; APPLICANT: Pyle, Ruth A.

```

APPLICANT: Indrias, Carol Joseph  
APPLICANT: lodes, Michael J.  
APPLICANT: Secrist, Heather

APPLICANT:	Fanger, Gary R.
APPLICANT:	Smith, Carole L.
APPLICANT:	Durham, Margarite

; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY  
 ; TITLE OF INVENTION: AND DIAGNOSIS OF COLON CANCER  
 ; FILE REFERENCE: 210121.563

```

; CURRENT FILING DATE: 2002-01-31
; NUMBER OF SEQ ID NOS: 3417
; SOFTWARE: FastSeq for Windows Version 4.0

```

```

; LENGTH: 477
;
; TYPE: DNA
; ORGANISM: Homo sapiens

```

Query Match	15.18;	Score 52.4;	DB 14;	Length 477;
Best Local Similarity	81.18;	Pred. No. 0.00047;		

4 ACATCAGTCTCTTCCAGAGTACCAACACTGTCAATGTGATGATGCCGCCCTCAGAGA 63

64 CCTCACCACGCGGACACATGCTTCGGTA 93  
 103 CCTGGCTTACGGGACACACGGCTCTGTA 7A

RESULT 12  
DS-10-066-543-2878/C

```
;
;
; sequence 2070, application US/1008003*
; publication No. US20030087818A1
; GENERAL INFORMATION:
; APPLICANT: Jiang Yuguin
```

APPLICANT: Xu, Jiangchun  
APPLICANT: Indrias, Carol yoseph

1       APPLICANT:       Lodes, Michael J  
 2       APPLICANT:       Secrist, Heather  
 3       APPLICANT:       Carter, Darlick  
 4       APPLICANT:       Fanger, Gary R.  
 5       APPLICANT:       Smith, Carole L.  
 6       APPLICANT:       Durham, Margalite  
 7       APPLICANT:       Stolk, John A.  
 8       TITLE OF INVENTION: COMPOSITION

FILE REFERENCE: 210121.563  
CURRENT APPLICATION NUMBER: US/10/066,543  
CURRENT FILING DATE: 2002-01-31

```
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 2878
LENGTH: 549
```

ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: misc\_feature

US-10-066-543-2878

A AACAATCACTGTCATTGGCCACAGTAACTAACACTGTCAATCTGCAGATTTATCATGCCCGCCCCCTCAGA 63  
Best Local Similarity 81.1%; Pred. NO. 0.0005;  
Matches 73; Conservative 0; Mismatches 16; Indels 1; Gaps 1.

Db 162 AACATCAGTCTTGGCCAGATTACGACACAGTCTACTTGACCATACTGA-CCCTCTTTA 108  
 54 CCTTCACCCATGGCGGACACAAATGCTTCCGGTA 93

Db 103 CCTGCCACGCGGACACACGCTCTCTGTA 74

US-10-066-543-2964/c  
; Sequence 2964, Application US/10066543  
; Publication No. US20030087818A1

```

; APPLICANT: Jiang, Yugu
; APPLICANT: Pyle, Ruth A.
; APPLICANT: Xu, Jiangchun

```

```

; APPLICANT: Iodes, Michael J.
; APPLICANT: Secrist, Heather
; APPLICANT: Carter, Darrick

```

;; APPLICANT: Smith, Carole L.  
;; APPLICANT: Durham, Margarita  
;; APPLICANT: Stolk, John A.

```

; TITLE OF INVENTION: AND DIAGNOSIS OF COLON CANCER
; FILE REFERENCE: 210121.563
; CURRENT APPLICATION NUMBER: US/10/066,543
;

```

```

; NUMBER OF SEQ ID NOS: 3417
;
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2964
; LENGTH: 567

```

```

; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature

```

US-10-066-543-2964

Query match 10.1% score 32.47 DB 14; length 307;  
Best Local Similarity 81.1% Pred. No. 0.00051;  
Matches 73; Conservative 0; Mismatches 16; Indels 1; Gaps 1

QY 4 AATATACAGTCTTGGCAAGTAAACAACACTCTCATGTATGTAAGCGCCCCCTTAGA 63  
 Db 162 AACATACAGTCTTGGCAAGTAAACAACACTCTCATGTATGTAAGCGCCCCCTTAGA 104  
 QY 64 CCTTACCCCAAGCGGACACATGCTTCCGGTA 93  
 Db 103 CCTGCGCCACAGCGGACACAGGCTCTCTGTA 74

```

RESULT 14
US-10-066-543-2864
Sequence 2864, Application US/10066543
Publication No. US20030087818A1
GENERAL INFORMATION:
APPLICANT: Jiang, Yugu
APPLICANT: Pye, Ruth A.
APPLICANT: Xu, Jiangchun
APPLICANT: Indrias, Carol Yoseph
APPLICANT: Lodes, Michael J.
APPLICANT: Secrist, Heather
APPLICANT: Carter, Darick
APPLICANT: Fanger, Gary R.
APPLICANT: Smith, Carole L.
APPLICANT: Durham, Margarla
APPLICANT: Stolk, John A.
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
TITLE OF INVENTION: AND DIAGNOSIS OF COLON CANCER
FILE REFERENCE: 210121.563
CURRENT APPLICATION NUMBER: US/10/066,543
CURRENT FILING DATE: 2002-01-31
NUMBER OF SEQ ID NOS: 3417
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 2864
LENGTH: 594
TYPE: DNA
ORGANISM: Homo sapiens
US-10-066-543-2864

```

[illegible]

```

RESULT 15
US-09-871-161-483/C
: Sequence 483: Application US/09871161
: Publication NO. US20030097666A1
: GENERAL INFORMATION:
: APPLICANT: ENEBE, WILSON O., ET AL.
: TITLE OF INVENTION: NOVEL HUMAN GENES AND GENE EXPRESSION
: FILE REFERENCE: CCDNA-260XX
: CURRENT APPLICATION NUMBER: US/09/871,161
: PRIORITY FILING DATE: 2001-05-31
: PRIOR APPLICATION NUMBER: 09/328,111
: PRIOR FILING DATE: 1999-06-08
: PRIOR APPLICATION NUMBER: 60/117,393
: PRIOR FILING DATE: 1999-01-27
: PRIOR APPLICATION NUMBER: 60/098,639
: PRIOR FILING DATE: 1998-08-31
: NUMBER OF SEQ ID NOS: 544
: SOFTWARE: PasteSeq for Windows Version 3.0
: SEQ ID NO 483
: LENGTH: 603

```

```

; TYPE: DNA
; ORGANISM: Homo sapiens
;
; FEATURE:
;
; NAME/KEY: misc_feature
; LOCATION: (1)..(605)
;
; OTHER INFORMATION: n = A,T,C or G
US-09-871-161-483

```

	Query Match	15.1%	Score 52.4;	DB 11;	Length 605;
	Best Local Similarity	81.1%;	Pred. NO. 0.0052;		
	Matches	73;	Conservative	0;	Mismatches 16;
					Indels 1;
					Gaps 1.
Oy	4	AAATCATGTCCTCCAGATACACACATGTCATGATTATGCCGCCCCCTCTAGA	63		
Db	162	AAATCATGTCCTCTCCAGATTACGACACATCTCTACTTGACCAATACTGA-CCCTCTTTA	104		
Oy	64	CCTCACCCACGCGGACACATGCTTCGGSTA	93		
Db	103	CCTGCGCCACGCGGACACACGCTCTCTGTA	74		

Search completed: August 26, 2003, 02:45:57  
Job time : 86.049 secs